The University of Osaka
Institutional Knowledge Archive

Title	複数のモダリティに基づく人間らしさの自動生成
Author(s)	境, くりま
Citation	大阪大学, 2017, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/61816
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

複数のモダリティに基づく
人間らしさの自動生成

平成 29 年 3 月

境 くりま

複数のモダリティに基づく
人間らしさの自動生成

博士（工学） 論文提出先
大阪大学大学院基礎工学研究科

平成 29 年 3 月

境 くりま

概　要

近年，ヒューマノイドロボットが社会的な役割を担う存在として注目されている．特に，人間に酷似したアンドロイドには雑用的な仕事よりも安心感や信頼性が必要となる仕事が適していると言われており，アンドロイドは，より信頼感が求められるような場面でも社会的役割を果たせる存在になると考えられる．そのような社会的存在となる上で，人間同士と同じ対話プロトコルをアフォードするアンドロイドでは，人間らしい動作を生成することが望まれる．また，身体動作は発話内容の理解の促進，感情理解の補助，発話の促進など様々な効果が報告されており，円滑な対話を実現する上でも人間らしい身体動作の実現は不可欠である．

アンドロイドの身体構造やアクチュエータの動特性は人間の筋骨格系とは異なる．そのようなハードウェアの制約があるため，アンドロイドで人間の動作を完全に再現することはできず，人間と同じ動作を行わせようとすると不自然さが残る．このように，人間と僅かに異なる動きや外見をしたものに対し，人間は恐怖を感じることが知られており（不気味の谷），アンドロイドとのインタラクションを悪化させてしまう．

以上の背景から，本研究では，僅かに足りていないアンドロイドの動きの人間らしさを補完し，不気味の谷を超え，対話の質を向上することを目標とする．しかし，上記したようにアンドロイドでは人間の動きの複製は困難であり，また写真上の動かないアンドロイドであっても，実際の人間ではないことがわかってしまう．そのため，外見や動きなどの単一モダリティを人間らしくするだけでは不十分である．しかし，人間は物質の質感を認識する際，視覚や聴覚といった複数の感覚モダリティの判断の論理積を計算するような形で，材質情報の統合が行われていることが報告されている．そのため，このような質感に類する人間らしさもマルチモーダルな感覚統合によって高めることができると考えられる．つまり，アンドロイドが複数のモダリティで人間らしさを同時に提示することで強く人間らしさを感じさせることが出来ると考える．

アンドロイドの振る舞いおいて，常に複数モダリティで人間らしさを同時に提示できるとは限らない．そのような問題に対して人間らしさを知覚させるのに重要なのは，提示のタイミングである．人間には一旦ある決断を行うと，その後に得られた情報を決断した内容に有利に解釈する傾向（確証バイアス）があり，一旦強く人間らしいと思い込ませることができれば，アンドロイドが再現できない動きがあったとしても，人間らしいという認知にバイアスされ，「人らしく動けるのにたまたま動かなかった」や「表情変化が少ない人」などの解釈をし，人間らしさを保持することができると考える．特に上記で述

べた動きと別モダリティの同期の仕方を誇張することで，断続的に強く人間らしさを印象づけられると考える．

本研究では，対話する上で必ずアンドロイドが行う発話と動きの同期性に着目し，発声に伴う症状的非言語行動，発話の意味と一致するジェスチャー，感情的発話と一致した症状的非言語行動について，発話に伴う人間らしい動作特徴を明らかにし，その特徴を誇張するように動作生成システムの構築した．

発声に伴う生理学的動作について，人間は声道や口の形を変え息を出すことで声を出すように，発声する韻律特徴に応じ発声しやすい姿勢があることを明らかにした．そのルールに従いアンドロイドの発話音声に基づいて，自動で体幹の動きを生成するシステムを構築し，自然な動作を生成することを心理実験により確認した．

発話の意味と一致するジェスチャーについて，アンドロイドを遠隔対話メディアとして利用し，操作者の感嘆詞発話に合わせ動作を自動で付加するシステムを構築した．感嘆詞発話に加え，感嘆詞動作を付加することで共感度合いを強めるなどインタラクションが向上することを心理実験により明らかにした．

感情的発話と一致した症状的非言語行動について，感情レベルで発話と一致した動作を生成するシステムを構築するためには，まずは感情的な動き方がどのような特徴になっているかを解明し，次に感情的な喋り方と感情的な動きをどのようなタイミングで同期させると効果的に人らしい感情を表現できるかを明らかにし，自動動作生成システムの構築を行う必要がある．本論文では，まず感情的な動き特徴を明らかにする段階まで行った．

目　次

第 1 章　序論　5

1.1　アンドロイドの意義と課題 5

1.2　アンドロイドの身体動作による対話の円滑化と課題 7

1.3　マルチモーダルの同期性に基づく動作生成 10

1.4　自律アンドロイドと遠隔操作型アンドロイドでの自動動作生成の検証　11

1.5　研究の位置づけと研究意義 12

第 2 章　関連研究　13

2.1　発声に伴う症状的非言語行動の自動生成システム 13

2.2　発話の意味と一致するジェスチャーの自動生成システム 15

2.3　発話時の感情に適した動作生成システム 16

2.4　遠隔操作型アンドロイドにおける動作生成システム 18

第 3 章　発声に伴う症状的非言語動作　21

3.1　まえがき 21

3.2　韻律と頭部動作の関係を見つける実験 21

3.2.1　実験設定 22

3.2.2　実験手順 22

3.2.3　実験結果 22

3.3　身体的拘束に基づく発話動作生成システム 24

3.3.1　2 次遅れ系フィルターを利用した頭部動作生成 25

3.3.2　韻律情報の抽出 26

3.3.3　首と腰の協調動作 27

3.3.4 動作データに基づくモデルのチューニング 27
3.4 評価実験 30
3.4.1 実験設定 30
3.4.2 評価指標 31
3.4.3 実験結果 32
3.5 考察 33
3.6 展望 36
3.7 まとめ 36

第 4 章 発話の意味と一致するジェスチャー 37
4.1 まえがき 37
4.2 頭部動作システムの構築 38
4.2.1 談話機能と頭部動作の関係性に関する知見 38
4.2.2 談話機能と発話の関係性に関する知見 39
4.2.3 言語情報の抽出 39
4.2.4 韻律情報の抽出 40
4.2.5 リアルタイム音声駆動頭部動作生成システム 41
4.3 提案システムの評価実験 44
4.3.1 実験目的 44
4.3.2 実験設定 44
4.3.3 実験システム 45
4.3.4 実験手順と評価指標 46
4.3.5 実験結果 47
4.4 考察 49
4.5 展望 51
4.6 まとめ 52

第 5 章 感情的発話と一致した症状的非言語動作 53
5.1 まえがき 53
5.2 実験設定 54
5.2.1 実験目的 54
5.2.2 実験システム 54

5.2.3 実験条件 55
5.2.4 実験手順 56
5.3 実験結果 57
5.4 考察 63
5.5 まとめ 66

第 6 章 結論 69

参考文献 75

謝 辞 87

第 1 章

序論

1.1　アンドロイドの意義と課題

ヒューマンロボットインタラクションの向上を目指し，様々なロボットが開発されてきた．人間がロボットとインタラクションを行う際，そのロボットの外見に基づいてそのロボットの能力や行いたいインタラクションを想定する [1]．特に，人間に酷似したアンドロイド（図 1.1）には雑用的な仕事よりも安心感や信頼性が必要となる仕事が適していると言われており [2]，アンドロイドは，より信頼感が求められるような場面でも社会的役割を果たせる存在になると考えられる．そのため，人間に酷似したロボット（アンドロイド）と人間を模したロボット（ヒューマノイドロボット）に対し，期待される役割は異なると考えられる．図 1.2 に示すように，ヒューマノイドロボットに対し，人は子供の面影を感じるような外見，または機械らしさを残すことで，人間側が立場が上の存在になるようデザインする傾向にある．一方で，アンドロイドのような人間と同じ外見をすることで，人間側が頼ることも可能な存在となり得る．例えば，アンドロイドのタスクとして，イベント会場の案内役 [3),4)]，デパートでの販売員 [5]，病院での陪席者 [6]，受付 [7] などの試みが行われている．そのように，人間の深層心理まで入る込めるアンドロイドと自然にインタラクションできるようになることで，ストレス社会に生きる現代人のカウンセラーなどにも利用できると考えられる．

その想定と実際のロボットの振る舞いの差異は適応ギャップ [8] と呼ばれる．ロボットに期待した振る舞いが実際に感じた振る舞いよりも高い場合は負の適応ギャップが生じ，ロボットに対する人の印象は悪くなる．一方で，実際の振る舞いが期待を超える場合は

図 1.1 アンドロイド

正の適応ギャップが生じ，ロボットに対する印象は良くなる．期待と実際の振る舞いが等しい場合は，人は思い通りのインタラクションを行える．ただし，この場合はロボットを単なる道具として認識している可能性が高い．インタラクティブなアンドロイドを設計する上では，アンドロイドに対する適応ギャップが負にならないように振る舞いを設計することが重要な課題となる．

船越らは機械らしさを残した外見のロボットを用いて人間の期待を下げ，そのロボットに適したロボットらしい表現 (Artificial Subtle Expression) を用いることで，人間の期待と実際のロボットの振る舞いとのギャップを小さくし，インタラクションの改善を試みている [9),10)]．しかし，人間の期待を下げるということは，人間がそのロボットに求めるインタラクションの次元を下げることになる [1)]．例えば，人間同士のインタラクションであっても，大人が子供相手に人生相談をしないのと同様に，対話相手によって適した対話内容を選ぶ．特に，介護や病院での陪席者などの社会的役割には，作業をこなす以上に安心感や信頼性といった精神的に支えることが出来る能力が必要であり，機械らしいロボットではそのような役割を果たせない．一方，同じアンドロイドであっても，人間のように状況に応じ外見を変えることが出来るため，アンドロイドは様々な社会的役割を果たすことが出来る．そのため，アンドロイドに焦点を当て，人間らしい外見から期待され

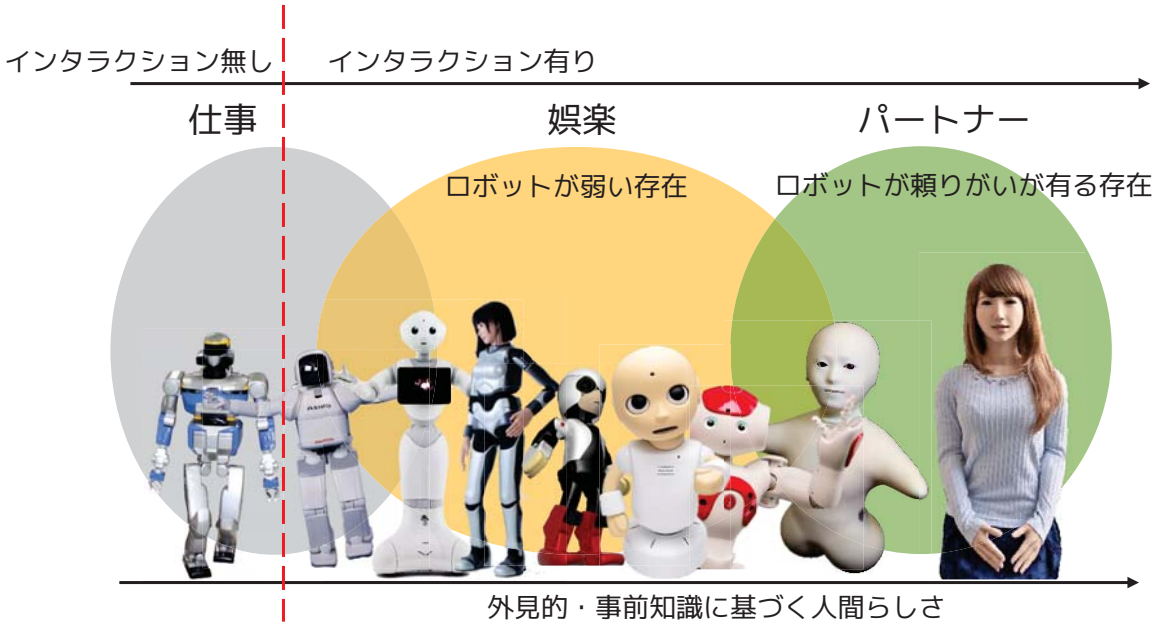


図 1.2 インタラクティブヒューマノイドロボット

る振る舞いの実装を行う．

1.2　アンドロイドの身体動作による対話の円滑化と課題

人間に酷似したアンドロイドに対し，インタラクションする対話者はアンドロイドに人間同士と同じインタラクションを期待する．人間同士の対話では，「ことばによらない」情報である非言語情報が重要である．これは以下の 3 つに分類される [11]．

周辺言語　音声情報から言葉の属性を除いたものすべて．例：音声を特徴づけるピッチ，抑揚，ストレス

身体動作　身体の動き．例：表情や身振り．手振り，姿勢

場　対話者との関係の中に成立する情報．例：対人距離，対人関係，空間の共有

さらに，身体動作は症状的非言語行動とジェスチャーに分類される [12]．症状的非言語行動とは，感情に応じた表情変化や発話時の口の動きなど，特に何かを表現しようという意図が直接の原因でなく起こるもので，情動や生理的内部状態が外から見てわかるような形で現れるものである．一方でジェスチャーは頷きやお辞儀など，あることを表現しよ

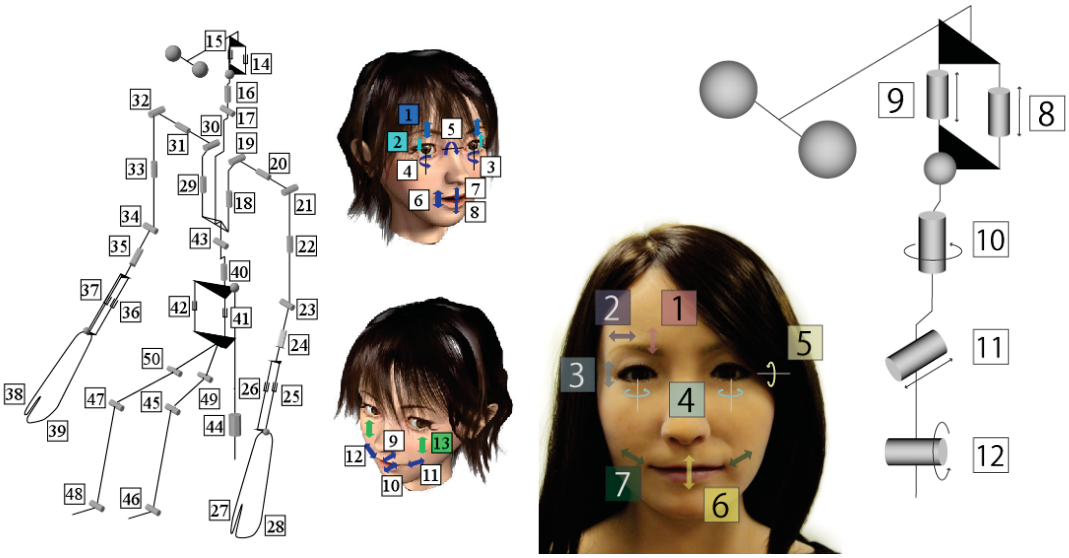
うという意図の達成に向けて生起する動きである．

そして，発話に含まれる言語的意味と動きなどの非言語情報が矛盾する場合では，対話する人は非言語情報を優先的に信用する傾向があり，非言語情報が円滑な対話に重要である（メラビアンの法則 [13]）．特に身体動作には，発話内容の理解の促進 [14–17]，感情理解や印象形成 [18–21]，発話や議論の促進 [22,23] といった効果が認められる．そのため，上記のような社会的存在となる上で，人間同士と同じ対話プロトコルをアフォードするアンドロイドでは，人間らしい身体動作を生成することが必要である．

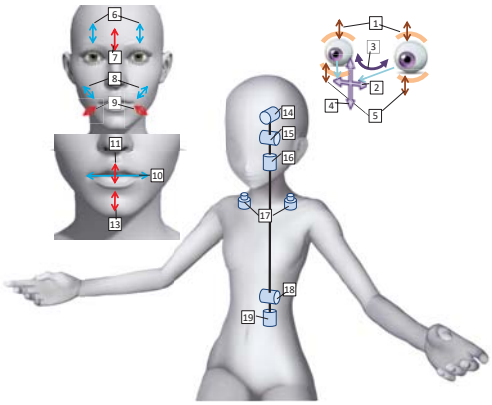
コンピューターグラフィックスの分野では，人間の動作をモーションキャプチャーシステムを用いて計測し，CG エージェントに計測動作を再現させることで人間らしい動作を実装している．歩行動作から投球動作を自然に連結させるといった異なる動作をつなぐ手法 [24] や CG エージェントの発話に合わせ自動で頭部動作 [25–31] や身振り手振り [32–34]，口唇動作 [35] を生成する手法が提案されている．これらの手法では計測した人間の動きを状況に合わせ再現することを目標にしている．しかし，アンドロイドの身体構造（図 1.3）やアクチュエータの動特性は人間の筋骨格系とは異なる．そのようなハードウェアの制約があるため，アンドロイドで人間の動作を完全に再現することはできず，人間と同じ動作を行わせようとすると不自然さが残る．このように，人間と僅かに異なる動きや外見をしたものに対し，人間は恐怖を感じることが知られており（不気味の谷 [36]），アンドロイドとのインタラクションを悪化させてしまう．

以上の背景から，本研究では，僅かに足りていないアンドロイドの動きの人間らしさを補完し，不気味の谷を超え，対話の質を向上することを目標とする．しかし，上記したようにアンドロイドでは人間の動きの複製は困難であり，また写真上の動かないアンドロイドであっても，実際の人間ではないことがわかってしまう [37]．そのため，外見や動きなどの単一モダリティを人間らしくするだけでは不十分である．一方，人間は物質の質感を認識する際，視覚や聴覚といった複数の感覚モダリティの判断の論理積を計算するような形で，材質情報の統合が行われていることが報告されている [38,39]．そのため，このような質感に類する人間らしさもマルチモーダルな感覚統合によって高めることができると考えられる．つまり，アンドロイドが複数のモダリティで人間らしさを同時に提示することで強く人間らしさを感じさせることが出来ると考えられる．

ただし，アンドロイドの振る舞いにおいて，常に複数モダリティで人間らしさを同時に提示できるとは限らない．そのような問題に対して人間らしさを知覚させるのに重要なのは，人間らしい動きの提示のタイミングである．人間には一旦ある決断を行うと，その



(a) GeminoidHI-2　　(b) GeminoidF



(c) ERICA

図 1.3 アンドロイドの関節配置

後に得られた情報を決断した内容に有利に解釈する傾向（確証バイアス [40]）がある．一旦強く人間らしいと思い込ませることができれば，アンドロイドが再現できない動きがあったとしても，人間らしいという認知にバイアスされ，「人らしく動けるのにたまたま動かなかった」や「表情変化が少ない人」などの解釈を誘発し，人間らしさの印象を保持

することができると考えられる．そこで，複数のどのようなモダリティが同期的に表出されると人間らしさを感じるのか，どのような表出方法が人間らしさのバイアスをもたらすのかを明らかにすることが重要である．石井ら[41]はアンドロイドの発話に合わせた口唇動作を生成する際，発話音声のフォルマント情報から母音を推定し，それに基づいて口の開き度合いを制御する手法を提案した．提案手法によって生成された口唇動作は，人の口唇動作をアンドロイドに複製した場合より，より自然な動きとなることが示されている．これは，母音にあった口唇動作が人間らしさには重要であることを示している．また，船山ら[42]は遠隔操作型アンドロイドの動作に，操作者の笑い声に合わせて典型的な笑い動作を付加することで，操作者自身の動作よりも自然な動作を生成できることを明らかにした．CG エージェントでは動きの加速度や振幅を誇張することで，人間らしさやアニメキャラクターらしさを表現できることがわかっている[43)–46]．これら動作生成の研究から，人間らしさを強く感じる特徴があり，その特徴を誇張することで制約のあるアンドロイドでも自然な動作と認識させることができると考えられる．特に上記で述べた動きモダリティと別のモダリティの同期の仕方を誇張することで，断続的に強く人間らしさを印象づけられると考えられる．また，人間同士が対話する場合でも，対話相手が見える時に比べ，見えない状態で対話する場合，つまり視覚モダリティが制限されたほうが，発話権を譲渡する際に生じる韻律特徴が誇張されることが知られている[47]．つまり，実際の人間もあるモダリティが制約されると別のモダリティを誇張することで円滑な意思疎通を図ろうとする．

以上をまとめると，**マルチモーダルの同期性，誇張による認知バイアスの誘発**がハードウェアの制約のあるアンドロイドで人間らしい動作を知覚させる鍵となる．

1.3　マルチモーダルの同期性に基づく動作生成

章 1.2 でも説明した通り，身体動作は症状的非言語行動とジェスチャーに分類される[12]．症状的非言語行動とは，感情に応じた表情変化や発話時の口の動きなど，特に何かを表現しようという意図が直接の原因でなく起こるもので，情動や生理的内部状態が外から見てわかるような形で現れるものである．一方でジェスチャーは頷きやお辞儀など，あることを表現しようという意図の達成に向けて生起する動きである．

行動決定の研究は多く行われており，visual saliency に基づく注視点の抽出[48]や複数人対話で誰を見るか[3]といった注視対象を選択する研究や，相槌の挿入タイミングを決

定する[49),50)]といったものがある．これら行動決定によりアンドロイドの発話文章と意味的ジェスチャが生成される．例えば，挨拶をする際には，「こんにちは」などの発話とお辞儀動作が同時に表出される．また，相槌のような反射的行動に対しても，「うんうん」などの発話と頷きが同期し生成される．このように**意味レベルでの発話とジェスチャーの同期**がある．

症状的非言語動作について，人間の発話音声の韻律特徴と頭部動作[51)]や体の動き[52)]には相関が高いことが報告されていように，発話と動きは意味より低いレベルにおいても同期が見られる．また，感情状態に応じて動きの速さや大きさが変わることも報告されている[53),54)]．このように，声を出すための動きという**生理的レベルでの発声と症状的非言語動作の同期**と，怒っている喋り方と動き方という**感情レベルでの発声と症状的非言語動作の同期**がある．

そこで，本研究では，対話する上で必ずアンドロイドが行う発話と動きの同期性に着目し，発声に伴う症状的非言語行動（第 3 章），発話の意味と一致するジェスチャー（第 4 章），感情的発話と一致した症状的非言語行動（第 5 章）について，発話に伴う人間らしい動作特徴を明らかにし，その特徴を誇張するように動作生成システムの構築を行う．

これらの評価実験結果から，多様なモダリティにおいて同期性がアンドロイドの人間らしさや自然さを向上させることが示す．

1.4　自律アンドロイドと遠隔操作型アンドロイドでの自動動作生成の検証

音声と同期する身体動作の自動生成技術は，生理学的動作から意味に伴う動作まで，すべての動作を自動で生成する自律アンドロイドでは必要な技術である．一方で，アンドロイドを遠隔操作し，コミュニケーションメディアとして用いる場合でもこの技術は有用である．遠隔操作では，操作者の動きをアンドロイドに写像することで生物らしい動きが生成できるはずである．しかし，遠隔操作ロボットを用いた対話においても，操作者側の対話インターフェースはビデオチャットと同様にモニタであり，モニタに映し出された相手を見ながら対話するようなインタフェースでは，対面時に比べて話者の頷きなどの社会的シグナル動作（意味的動作）が少なくなる[55)]．そのため，意味的動作の自動生成は自律アンドロイドのみならず，遠隔操作型アンドロイドを用いた遠隔対話の向上にも役立つと考えられる．しかし，遠隔操作の場合では，操作者本人の動作とは異なる動

作を自動的に付加することで，対話者は違和感を感じる可能性がある．そこで，遠隔操作型アンドロイドでの意味的動作の自動生成の有効性については第 4 章で検証する．

1.5　研究の位置づけと研究意義

本研究の位置づけは，人間と同じ動きの自動生成ではなく，アンドロイドを人間と知覚させるための動作生成手法を明らかにすることである．本論文では，アンドロイドが出力する音声と同期した動作を自動で生成することで，マルチモーダルな感覚統合によって人間らしい質感を高める手法を明らかにする．さらに動作生成システムを構築する上で，自律アンドロイド・遠隔操作コミュニケーションの両方で利用できるシステムの開発を目指す．また，人間らしい動きは，外見にかかわらず人型エージェントに対する親密度を向上させることが報告されており [56]，人間らしい動きを感じさせる要因を明らかにすることは，人型エージェント全般において意義がある．本研究を利用することで，従来のヒューマノイドロボットの行動生成システムを，アンドロイドに移植する際に，自動的に人間らしい動作に変換することができるようになる．

第 2 章

関連研究

第 2 章では，人間の発話と動作の同期性およびそれを利用した CG エージェントやロボットの動作生成に関する従来研究について，症状的非言語行動レベルと意味的行動レベルで説明する．また，第 5 章に関連して，CG エージェントやロボットの感情を表現する動作生成手法に関する従来研究，生理学的に感情が動作にどのように影響するかを調べた従来研究についてもこの章で述べる．さらに，第 4 章に関連して，遠隔操作型ロボットの動作生成に関する従来手法を紹介し，遠隔操作型ロボットに動作を自動的に付加する際の問題点について述べる．

2.1　発声に伴う症状的非言語行動の自動生成システム

コンピューターグラフィックスの研究分野では，エージェントの発話に合わせ頭部動作を自動生成する手法がいくつか提案されている．Le et.al. は発話音声のパワー，ピッチと頭部の 3 自由度の動きの関係を Gaussian Mixture Model を用いてモデル化し，リアルタイムで頭部動作を生成するシステムを提案している [29]．また，隠れマルコフモデルを用いた同様のモデル化も行われている [26)–28)]．しかし機械学習を用いた自動生成システムでは，学習に使われているモーションデータが収録された状況に合った動作しか生成できない．特に話す動作は対話相手との関係性により変化するため，あらゆる状況をあらかじめ想定し動作を記録して学習することは困難である．また，これら手法は収録されたデータを復元することを目的にしているため，異なる状況で使用するための動きの変調や他の動きと複合することができない．一方で，対話状況に合わせ複数の動作をミキ

シングする方法[57)–59)]の方が，様々な状況に適応できる．そのため，どのような状況においても共通に現れる動作である発声のためだけの動きを生成するシステムを構築することが有効である．さらに，動きがパラメータ化された発声動作生成モデルを構築することで，状況に合わせた動きの調節が期待される．

アンドロイドの身体構造やアクチュエータの動特性は人間の筋骨格系とは異なるため，アンドロイドで人間の動作を完全に再現することはできず，人間と同じ動作を行わせようとすると不自然さが残る．また，ヒューマンフィギュアにおける，主観的に同じ姿勢と感じるのは，関節角度の一致ではなく，手先位置の一致が重要であることが報告されている[60)]．そのため，アンドロイドのハードウェア的な拘束や，主観的人らしいさの認識が関節角の一致ではないため，上記の人間の関節角度を再現しようとする機械学習を用いた動作生成手法ではアンドロイドに適した動作生成ができない．

発声のみに伴う動きを生成する研究として，石井ら[41)]の口唇動作生成の研究，船山ら[42)]の笑いの典型的な動きの生成に関する研究がある．石井ら[41)]はアンドロイドの発話に合わせた口唇動作を生成する際，発話音声のフォルマント情報から母音を推定し，それに基づいて口の開き度合いを制御する手法を提案し，人の口唇動作をアンドロイドに複製した場合より，より自然な動きとなることを示した．船山ら[42)]は遠隔操作型アンドロイドの動作に，操作者の笑い声に合わせて典型的な笑い動作を付加することで，操作者自身の動作よりも自然な動作を生成できることを明らかにした．しかし，石井ら[41)]の口唇動作生成システムでは口唇動作しか生成できず，また，船山ら[42)]の笑い動作生成システムでは笑っている間の動きしか生成できず，発声のためだけの頭部動作はモデル化されていない．

発声に伴う動きそのものではなく，動きのタイミングを生成する研究もおこなわれている．Watanabe et.al. は，発話の on/off 情報から頷きのタイミングを推定する手法を提案している[61)]．しかし，頷き生成のタイミングを生成するだけで，どのような関節の動きが人間らしさを生むかまでわかっておらず，実際のアンドロイドで使用するには不十分である．

一方で，解剖学の知見から，口の開閉動作に伴い頭部が動くことも報告されている[62)]．図 2.1 に示すように，下顎の位置を変えずに顔を上方向にそらすことで口を開く場合と，下顎の開きに伴い顔が下方向に向く場合がある．この知見から頭部の発話動作も社会的状況の要素以外の身体的拘束をもとに生成できる可能性がある．しかし，この実験は発声を伴わない口と頭部動作の関係を調べたものであるため，発声を行う際の口の動きと

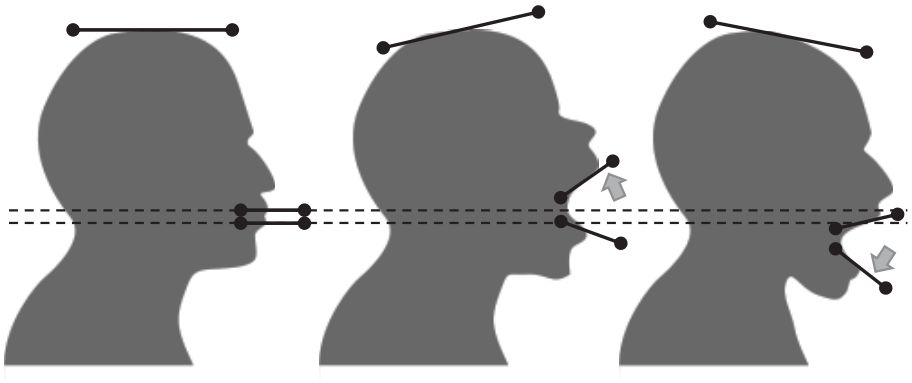


図 2.1 口の開閉に伴う頭部動作の変化

頭部動作の動きは調べられていない．

2.2　発話の意味と一致するジェスチャーの自動生成システム

意味的動作とは，相槌や指差しやお辞儀などの動作が情報伝達を果たすものを指し，人間同士の対話では重要な役割をしている．特に頭部動作は重要であり，相槌は発話や議論を促進する効果 [22),23)] があり，驚きや感心を表す頭部を上げる動作を伴うことで感情理解や印象形成が正確になることが報告されている [18)–21)]．

このような動作では，発話と動作の間に強い同期性が存在する．例えば，相槌や肯定の際には「はい」と言いながら頷く．このような強い関係性を用いれば，頭部の動作（例：頷き）を発話内容（例：「はい」という音声）から推定できる．従来研究より，発話の意味を表す談話機能と幾つかの頭部動作の同期性が報告されており [63)]，その同期性に基づく頭部動作モデルが提案されている [64),65)]．発話者の意図を表現するうえで，この動作生成手法では音声と動作で同じ意図を冗長に伝達している (例: 「はい」と言いながら頷く) が，実験結果では冗長であっても音声と動作を同時に伝えることによりロボットの自然さが向上したことが示されている．しかし，この評価実験では，談話機能の識別が手動で行われており，自動的に動作を生成できるシステムは構築されておらず，実際のインタラクションへの影響が検証されていない．

2.1 節で述べた機械学習を用いた頭部動作を生成するシステム [26)–29)] では，韻律特徴だけを入力とするため発話の意味と一致した動作の生成はできない．Watanabe et al.[61),66)] は，音声の ON-OFF 情報を入力とした頷き動作の予測モデルを構築し，コンピュータグ

ラフィックスのエージェントやヒューマノイドロボットの頷き動作をリアルタイムで生成するシステムを考案している．Wu et al.[67] は，対話エージェントの発話文章に含まれる特定のキーワードに合わせ，正弦波を用いた頷き動作を自動生成するシステムを提案している．Watanabe et al.[61),66] の手法も Wu et al.[67] の手法も，頷き動作しか生成できず，発話者の否定の意味，困惑，驚きなどを表現することができない．

2.3　発話時の感情に適した動作生成システム

感情を表現する動作生成に関しては従来から多くの研究が行われている．例えば，Miwa et al.[68] は，ニュートラル，苛立ち，不安，悲しい，幸せ，驚き，怒りを表す自由度を持つヒューマノイドロボット WE-4RII を開発した．また，Nakano and Hoshino[69] は，内部状態に対応するしぐさを選択することでエージェントの心理状態を表出する手法を提案している．Miwa et al.[68] の手法は離散的な感情空間を使用するものであり，また，Nakano and Hoshino[69] の手法では感情空間は連続的であるが，動作は予め定義されたシンボリックな動きを扱っている．そのため，感情ごとに動きを事前に用意する必要があり，複雑な感情が増えるほど動作を用意する手間が増え，感情の細かな変化を動作の細かな変化で表現できる手法にはなっていない．

感情的な動作をシンボリックに定義しない方法としては，機械学習を用いて発話音声の韻律特徴から発話動作を生成する試みがある [26)–29)]．しかし機械学習を用いたシステムでは，学習に使われているモーションデータが収録された状況に合った動作しか生成できない．学習データに含まれる動きを再現する手法で多様な状況における動きの変化に対応するためには，多様な状況での学習データを集める必要があるが，話し方は対話相手との関係性によっても変化するため，多様な状況に対応できるためのデータを集めることは困難である．したがって，学習データの動きを再現するような手法は，感情や態度の変化に適した動作を生成するシステムには適さないと考えられる．一方，著者らが第 3 章にて提案する頭部動作生成システム [70] は，人が発話する際の発声を補助する動きを，力学モデルを用いて生成するもので，発話情報に基づいてそれと同期した頭部動作を生成する．モデルのパラメータを変更することで，音声との同期性を保持しつつも，その動き方（動きの大きさ（振幅）や動きの速さ（速度）など）を連続的に変更することが可能である．しかしながら，感情に適した動作変調のモデルはまだ提案されていない．

動作を変調することで感情状態を表現する従来研究もある．人間の感情は苛立ち，不

安，悲しい，幸せ，驚き，怒りなど明確に分類できるわけではなく，複数の連続的な次元で表現できると考えられている．Russell[71] は，快–不快と覚醒–眠気を軸とする 2 次元上に様々な感情が配置される円環モデルを提案している．このような連続的な感情空間と動作を関連づける研究が行われている．Jia et al.[72] は，PAD モデル（Russell の円環モデルの拡張版）に合わせて，発話文章中の強調語（とても，すごくなど）に伴う頷き動作の振幅を変調する動作生成システムを提案している．この手法は楽しいなどのポジティブな感情のみを扱っており，感情空間全域での頷き動作と感情の関係は明らかにされてはいない．また，頭部の頷き動作を正弦波運動で表現し，その振幅と感情の関係を機械学習によって構築するため，全身の動きや頷き以外の動作でも同様の関係性が成り立つことを期待できる表現にはなっていないと考えられる．バイオロジカルモーション [73] を用いた研究では，悲観的な感情状態では動き（歩行動作）の振幅が小さくなり，速度が落ちることが明らかにされている [74]．また，ラバン理論 [75] とロボットジェスチャを組み合わせることで，Russell の円環モデルに合わせてジェスチャを変化させる手法が提案されている [76]．ただし，この手法で扱う動作は，ロボット用に誇張されたジェスチャ動作であり，人に酷似したアンドロイドの自然な動きの生成は期待できない．さらに，歩行時のバイオロジカルモーションからラバン特徴を抽出し，感情との関係を調べた研究 [54] や，怒りと悲しみの感情状態での人の蹴り上げ動作の変化を比較した研究 [53] もある．これらの研究から共通して示唆されることは，動きの大きさや速さが感情によって変化するということである．

心理状態と動き方の関係については，人間の生理学的な側面からも明らかにされている．中 [77] は，心理的な負荷が筋肉の弾性を高めることを明らかにしている．また宇尾野 [78] は，交感神経の働きにより骨格筋が緊張・弛緩することを，さらに山下 [79] は，感情が交感神経・副交感神経活動を活発にすることを明らかにしている．ラバン身体動作表現理論 [75] も基本的には，筋肉の緊張・弛緩度合いが心理状態を表している．例えば Nakano and Hoshino[69] は，リラックスした状態では胴体の動きが緩やかになり，反対に緊張した状態ではぎごちなくなることを明らかにしている．以上の生理学的な知見から，感情が筋肉の緊張・弛緩に影響を及ぼすことで，感情に応じて人の動き方が変化すると考えられる．

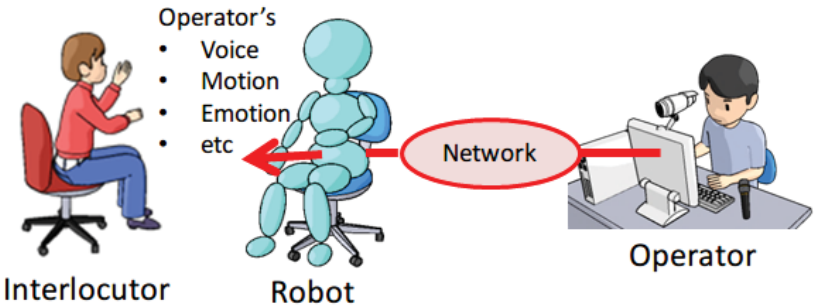


図 2.2 遠隔操作型アンドロイドを用いた遠隔対話

2.4　遠隔操作型アンドロイドにおける動作生成システム

近年，遠隔操作ロボットをコミュニケーションメディア (ロボットアバター) として用いることで，それらの非言語情報の伝達が試みられている (図 2.2)[80)–83)]．ロボットが操作者と同期して発話，動作することで，ロボットと対面する人はあたかも操作者が目の前にいるように感じ，遠隔地にいる人と円滑に対話できる [84)]．

従来の遠隔操作ロボットの動作生成手法は，操作者の動きを計測し，ロボットと操作者の関節角度の類似性に基づいて，操作者の動きをロボットに写像する手法が一般的である [81),82),85)]．遠隔操作ロボットを用いた対話においても，操作者側の対話インターフェースはビデオチャットと同様にモニタである．モニタに映し出された相手を見ながら対話するようなインタフェースでは，対面時に比べて話者の頷きなどの動作が少なくなる傾向にあることが報告されている [55)]．そのため，操作者の動作がそのまま写像されたロボットの動作は対面時のような自然な動作には至らず，ロボットアバターの効果が十分に発揮できない．

中道らは，遠隔操作インタフェースを用いた状態でも，操作者が対面時のような動作を行えるようにするための訓練手法を提案している [86)]．しかし，単なる対話のために訓練を要することは，ユーザーにとって余分な負荷となる．そこで，本論文で提案する音声から動作を生成する手法を遠隔操作で用いると，遠隔操作インタフェースによって減少した操作者の動作を自動的に補償し，遠隔対話を円滑にすることができる．

Tamaki et al.[55)] によれば，ビデオチャットインタフェースによって，頷きなどの意図（同意や肯定など）を表す動作が対面対話と比較して抑制されることが明らかにされている．ただし，同意や肯定などの意図を伝える回数自体が減少しているわけではない．す

なわち，ビデオチャットインタフェースでは，動作を伴う相槌が減り，音声のみの意図伝達が増えることが明らかにされている．従って，音声から自動生成した動作を遠隔操作アンドロイドに付加しても問題ないと考えられる．しかし，操作者にとっては意図していない動きが自動的に付加されていることで対話に違和感を覚える可能性もあるため，自動付加が遠隔操作における有効性の検証が必要である．

第 3 章

発声に伴う症状的非言語動作

3.1　まえがき

第 3 章では，社会的状況に依存せず，純粋に発話のための動作を，人間の身体的拘束を利用し発話情報に基づいてリアルタイムで生成することを目的とする．機械学習で構築したモデルでは，発話と動作のどのような特徴が人間らしさに関わっているのか解析しにくく，対話状況や発話者個性に合わせ動作を変調することも難しい．本研究では，発話と動作の関係がわかりやすく，そして動作を変調しやすいような動作生成モデルの構築を目指す．特に，目線をそらす動作は対話のコンテキストに依存し [87]，そのパターンは個性に依存する [88] ことから，本章では発話に合わせた首と腰の縦方向の動きに着目する．人間と同じ動きができないアンドロイドで，人間らしいと知覚させる鍵となる方法は，生理学的な発話と身体の縦方向の動きの同期を誇張し表出することである．

3.2　韻律と頭部動作の関係を見つける実験

本節では人間らしい発話動作を自動生成するためのルールを見つけるための実験を説明する．人間が発声する際頭部動作などが音声に同期することが報告されており，特にパワーとピッチの変化と動作の変化が同期することが知られている [52]．しかし，日本語ではパワー，ピッチの韻律特徴と頭部動作の相関は高くないことも報告されている [89]．また，解剖学の知見から，口の開閉動作に伴い頭部が動くことも報告されている [62]．そこで，音声のパワー，ピッチ特徴に加え，口の開き度合いの 3 特徴が，他人とのインタラクションを伴わない発声時に，動きとどのような関係があるのか，人の発声動作を観察す

ることで明らかにする．

3.2.1　実験設定

音声のパワー，ピッチと頭部動作の関係，口の開き度合いと頭部動作の関係を観察する実験を行った．口の開閉が母音を発音する際に大きく変化するため，実験参加者に「あ・い・う・え・お」を 3 秒間発声してもらい，その発声に伴う首の動きの変化を計測した．母音の発声はそれぞれを高音・中音・低音で発音する条件 (Voice Pitch Condition) と，発声しやすい声の高さで大声で発声する条件 (Mouth Openness Condition) を設けた．発声の種類は「高い “あ”」「低い “い”」のように指示をした．被験者には，各発声ごとに正面を一旦向くよう指示を出し，姿勢をリセットした．予備実験より，被験者は母音を発声する際に 2 要因（高音で大きな声など）を混同させると発声しづらかったため，本実験では，2 要因を分けて頭部動作の変化を計測した．また，小さな声で発声すると頭部が動かないことも予備実験にて確認されていたため，Mouth Openness Condition では，大きな声のみ発声させた．

頭部動作は被験者の頭頂に取り付けた Inertial Measurement Unit(IMU) で計測した．被験者には口の形をはっきり作るように教示することで，母音に対する口の開き具合を統制した．

3.2.2　実験手順

各条件ごとに被験者には 2 回試行させた．1 回目は実験室での発声に馴化するために行った．また，身体動作を正しく計測できているかの確認も行った．

3.2.3　実験結果

実験被験者は 11 人（男 : 6 人，女 : 5 人，平均年齢 22.0，標準分散 0.54）であった．そのうち男性被験者 1 人が正しく声の高さを発声できていなかったため解析から除いた．

Voice Pitch Condition の計測結果を図 3.1 に示す．縦軸は発声定常状態での首の角度を示す．高音，中音，低音を発声する際の首の角度を分散分析にかけたところ，有意差が認められた ($F(2, 18) = 12.843, p < 0.01$)．さらに，多重比較したところ，高音を発声する際に首の角度が最も上がり ($p < 0.05$)，低音を発声する際に最も下がることが明らかとなっ

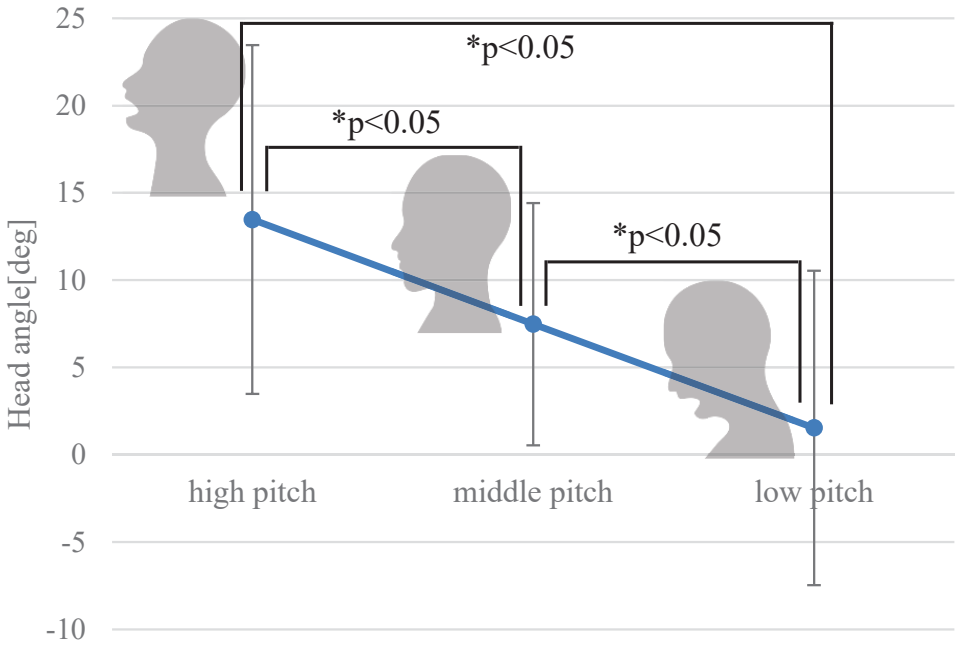


図 3.1 音程ごとの頭部姿勢

た ($p < 0.05$)．すなわち，高音を発声する際は頭部をそらし，低音を発声する際は頭部を下げる傾向が認められた．

Mouth Openness Condition の計測結果を図 3.2 に示す．縦軸は発声に伴う首の角度の変化量を示す．この変化量は，発話開始前と発声定常状態での首の角度の差の絶対値で定義した．口を開いて発声する「あ」「え」「お」群と口を閉じて発声する「い」「う」群に分け，発声に伴う首の角度の変化量の大きさを比較したところ，口の開きを伴う発声条件のほうが有意に首を大きく動かすことが認められた (ウィルコクソンの順位和検定，$p < 0.05$)．

以下にアンケートによる発声しやすい姿勢についての自由記述結果を示す．この記述からも，高音を発声する際は頭部をそらし，低音を発声する際は頭部を下げる傾向が認められた．

- 声の高低を意識して使い分けることが難しく感じ、高く出そうと思えば背筋が伸び顎が上がりました。低く出そうと思えば、背筋を少しだけ丸め顎を引き、なるべく口の中に篭るように発声しました。

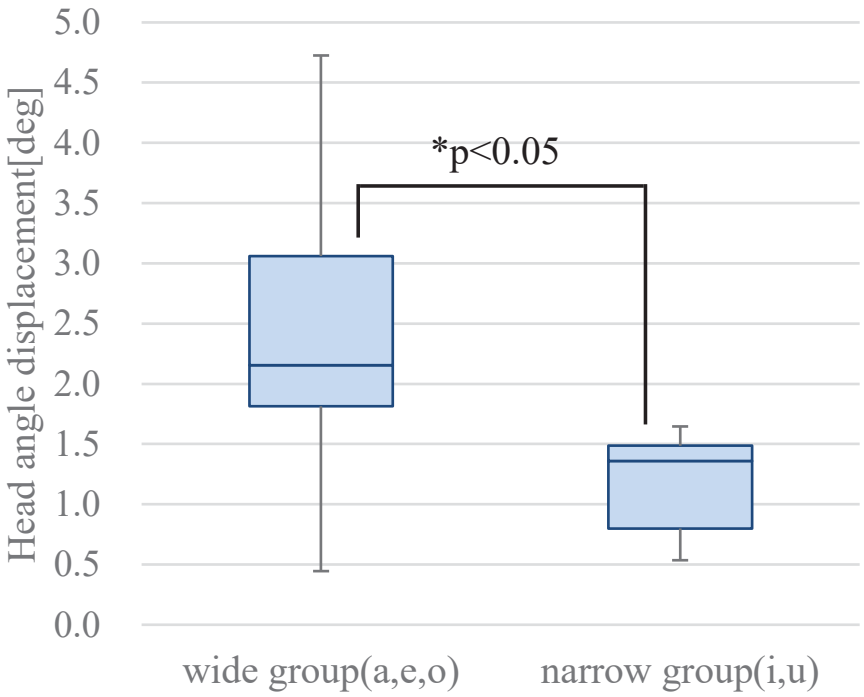


図 3.2 口の開閉度合いに応じた頭部の動作変化量

- 高い音を出す際は上を向き、低い音を出す際には下を向く
- 口を大きく開ける あといは上から声を出し、うえおは下からあげるイメージで声を出す 体の中心に力を集めるイメージ
- 高い音は背筋が伸びる感じでした。低い音になるほど下を向いていたと思います。
- 高い音を出すときは顔を上向きに、逆に低い音を出すときは下向きにすると出しやすかった

3.3　身体的拘束に基づく発話動作生成システム

以上の知見をもとに音声特徴から頭部動作を生成するアルゴリズムを以下に説明する．人間らしい動作には滑らかな関節制御が重要である [56),90)]．しかし，ピッチや口の開閉は急激に変化することがある．そのため，急激な変化の入力に対して，滑らかな運動を出力するフィルターが必要である．2 次遅れ系のダイナミクスに基づいて生成される動作が人間らしさ印象を与えることが報告されている [91)]．また，2 次遅れ系であるバネ–ダンパ系

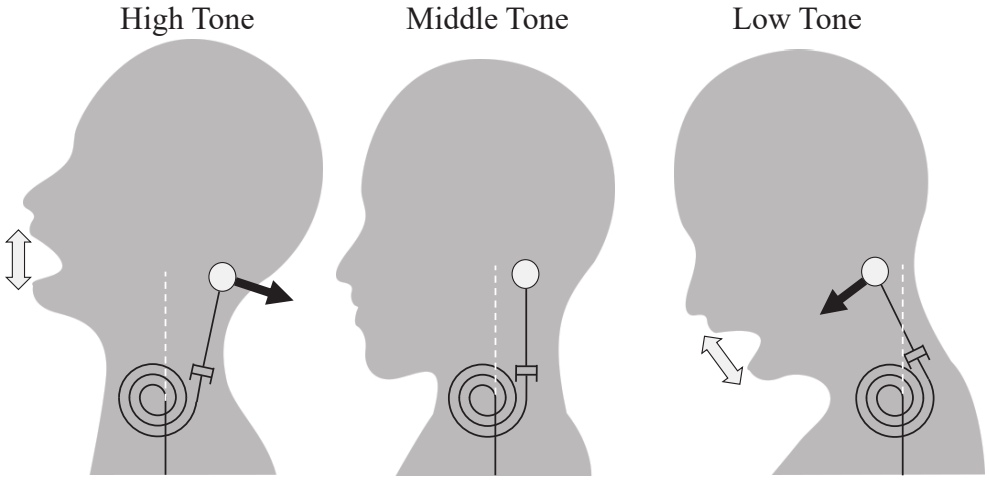


図 3.3 2 次遅れ系フィルターを利用した頭部動作生成

を用いた筋肉のモデル化が行われている [92),93)]．そのため，2 次遅れ系のフィルターを用いることで，間欠的な韻律特徴から人らしい動きに変換することができ，さらに，緊張度合・感情状態によって変化する筋肉の硬さに応じた動きの変調も，筋肉の動特性に類似した 2 次遅れ系のフィルターでは実現できると考えられる．

3.3.1　2 次遅れ系フィルターを利用した頭部動作生成

前節までの知見に基づき，音声の大きさ，ピッチおよび口の開閉度に基づいて，首の角度を制御する．口を大きく開けると首も大きく動き，また予備実験から大きな声を出さないと首が顕著に動かなかったことから，時刻 t における首を動かす大きさ（角度絶対値）$T(t)$ は口の開口度合と声の大きさに比例するようにする．また，高音域を発声する場合は頭部をそらし，低音域を発声する場合は頷く方向に首を動かし，中音域では首を動かさない傾向があることから，時刻 t における首の動く方向 $Dir(t)$（$= \pm 1$）は声の高さで決定する．ただし，説明したように，首角度をこのように定めると，急激な音声特徴の変化がある場合に首の動きが不自然になるため，前節で説明したように 2 次系のダイナミクスで追従するようにする．首の角度を θ_{base} とすると，下記のように計算される．

$$J\ddot{\theta}_{base} + D\dot{\theta}_{base} + K\theta_{base} = T(t)Dir(t) \tag{3.3.1}$$

首の動きの大きさを決める $T(t)$ は，口の開口度合に比例する $M(t)$ と，声の大きさに

比例する $P(t)$ の和で定義する（式 3.3.2）．V と L は声の大きさと口の開き度合という異なるスケールの外力を合わせるための定数である．口の開口度合に比例する $M(t)$ は式 3.3.3 のように定義する．口の開きが大きくまたは均一である場合は，$M(t)$ は口の開きの大きさ $DoM(t)$ に比例するようにする．口の開きが小さくなる場合は，$M(t)$ をなくすことでフィルターにより基準位置へ滑らかに戻る．また，声の大きさに比例する $P(t)$ は式 3.3.4 のように定義する．口の開き度合同様に，声のパワーが増えるまたは均一である場合は，$P(t)$ は声の大きさ $Power(t)$ に比例するようにする．声が小さくなる場合は，$P(t)$ をなくすことでフィルターより基準位置へ滑らかに戻る．口の開閉度合（$DoM(t)$）は Ishi et.al. のフォルマント抽出に基づく口唇動作推定の手法を用いる [41]．

$$T(t) = VP(t) + LM(t) \tag{3.3.2}$$

$$M(t) = \begin{cases} DoM(t) & (DoM(t) >= DoM(t-1)) \\ 0 & (otherwise) \end{cases} \tag{3.3.3}$$

$$P(t) = \begin{cases} Power(t) & (Power(t) >= Power(t-1)) \\ 0 & (otherwise) \end{cases} \tag{3.3.4}$$

首の動く方向を決定する $Dir(t)$ は声の高さに基づき決定する（式 3.3.5）．式 3.3.5 は，高音域を発声する場合は頭部をそらし，低音域を発声する場合は頷く方向に首を動かし，中音域では首を動かさないことを表す．声の高さの識別は次節で説明する．

$$Dir(t) = \begin{cases} 1 & (Headup) & (pitch = High) \\ -1 & (Headdown) & (pitch = Low) \\ 0 & (Restoringmovement) & (pitch = Middle) \end{cases} \tag{3.3.5}$$

頭部と頸部筋肉からなる系のダイナミクスをモデル化し，そのモデルから発話特徴に従って頭部動作を生成する方法も考えられるが，話者ごとやロボットごとに系の次数も変わる可能性がある．本論文では，特定の話者の動きの再現を考えているのではなく，また，構築するシステムがロボットに依存せずに使えることも想定するため，上記のようなアプローチで動作生成システムを構築する．

3.3.2　韻律情報の抽出

基本周波数 (F0) の値の抽出には，32 ms のフレーム幅で 10 ms 毎に LPC(Lear Predictive Coding) 逆フィルタによる残差波形の自己相関関数の最大ピークに基づいた処理を行う．

さらに，人間のイントネーションの知覚特性と一致するよう，F0 の値を対数スケールに変換した．

$$F0[\text{semitone}] = 12\log_2(F0[\text{Hz}]) \tag{3.3.6}$$

次に，最新の 100msec 間の $F0$ の平均値を計算し $\overline{F0}$ とした．そして，音調は式 4.2.3 に応じて，高音域，低音域，中音域に分類した．

$$pitch = \begin{cases} High & \overline{F0} > F0_{high} \\ Low & \overline{F0} < F0_{low} \\ Middle & (\text{otherwise}) \end{cases} \tag{3.3.7}$$

3.3.3　首と腰の協調動作

頭部が動く際には上下方向だけではなく，前後方向にも動くことが判っている [94]．このことから，首の 1 自由度の回転だけではなく，腰も連動させることでより人間らしい動きが実現できると考えられる．また，口と首の動き出すタイミングは異なり，口のほうがやや早いく動くことが報告されていることから [95]，動かす関節により位相差があることが考えられる．そこで，式 3.3.8 の変換式を用いて図 3.4 のような協調動作を実装する．act_i はロボットのアクチュエータを指し，α_{act_i}，β_{act_i} は各アクチュエータごとに設定する必要がある．

$$\theta_{act_i}(t) = \alpha_{act_i}\theta_{base}(t + \beta_{act_i}) \tag{3.3.8}$$

3.3.4　動作データに基づくモデルのチューニング

提案モデルと実際の人間の動きを比較するために，人間の発話音声と動きを録画し比較した．読み上げる文章は，ERICA(図 3.5) の自己紹介文章を用いた．動作データは Inertial Measurement Unit(IMU) を頭部と胴体の 2 箇所につけ，首と腰の動きを計測した．また，発話の仕方に個性がある可能性があるため，2 人の女性（話者 M と話者 H）のデータを収録した．

提案手法を用いてアンドロイドの頭部動作を行った．パラメータ J, D は 0.0676 と 0.52 に設定した．$\theta_{base} >= 0$ の場合は重力を加味し，K は 0.195($\theta_{base} \geq 0$)，0.065 ($\theta_{base} < 0$) に設定した．声の音圧はおおよそ 10 dB から 20 dB であり，口の開き度合いは 0 から 200（無単位）であることから，式 3.3.2 の V と L は 0.001 と 0.0005 に設定した．式 4.2.3 の

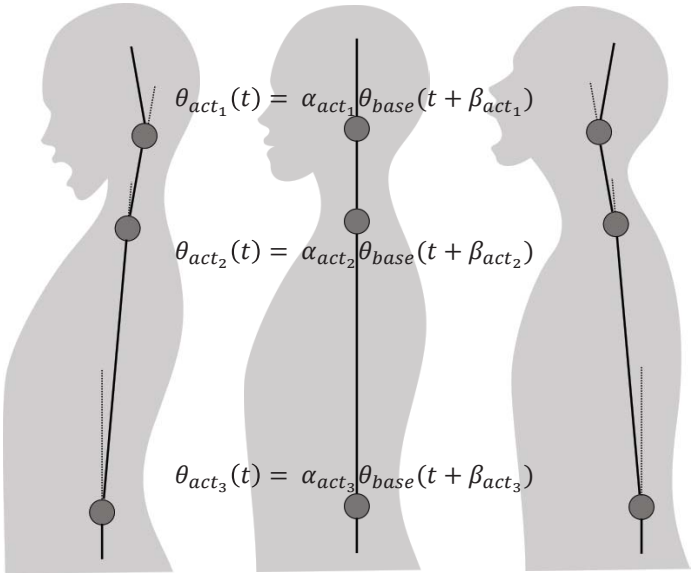
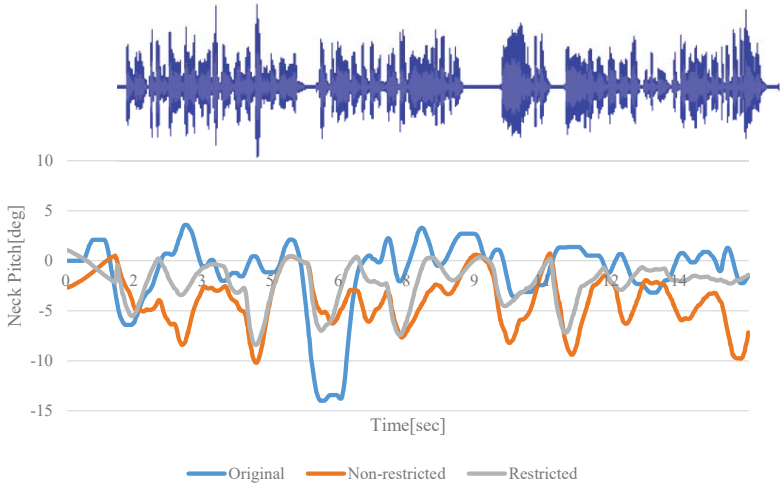


図 3.4 首と腰の協調動作

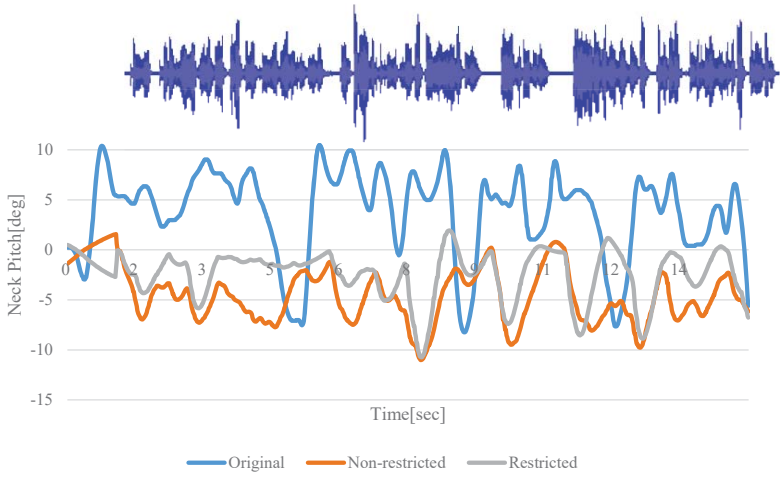
図 3.5 アンドロイド ERICA[96)]

閾値 $F0_{high}$ と $F0_{low}$ は 256 Hz と 215 Hz に設定した．これらパラメータは発話に合うように予備実験から実験者が設定した．口の開閉度合は Ishi et.al. のフォルマント抽出に基づく口唇動作推定の手法を用いた [41)]．算出された θ_{base} はアンドロイドの首の制御にマッピングされ (i.e., $\theta_{neck}(t) = \theta_{base}(t)$)，他の体幹は制御していない．

収録した音声データを用いて上記のモデルを用いて動作生成したものを図 3.6 に示す．



(a) 話者 H



(b) 話者 M

図 3.6 計測された頭部動作

Original は収録した実際の人間の関節角を示し，上記のモデルを用いて生成した関節角 ($\theta_{neck}(t)$) が Non-restricted である．図 3.6(b) では，Original ではほとんど首の角度が正である（話者 M が発話中上方を向いている）．一方で，提案手法を用いた方は首の角度がほ

とんど負である．本研究では，首の角度と韻律特徴の相関に焦点を当てているため，話者がどこを向いているかは考慮しない. Original と Non-restricted を比較すると，Original は微小な動きをする定常状態と，時折大きな動きの組み合わせになっている．一方，提案手法である Non-restricted では常に同じような振幅，周期の動きになっている．単一の周期的な動きのみだと，対話する人がそのパターンに気づき，ロボットらしさを帰属させる可能性がある．そこで，Original のような大きな動きと小さな定常状態の動きを表現するために，上記で提案した式 3.3.4,3.3.3 のモデルに閾値を設け，動作生成の頻度を軽減したモデルを提案する（式 3.3.9,3.3.10）．このモデルでは，閾値を超えた際に大きな動きが生成され，それ以外では 2 次遅れ系フィルターにより微小振動を行うようになる．このモデルを用いて生成した動作が Restricted である．図 3.6 の Restricted がこのモデルにより生成された動作である. 今回は閾値 P,M は 1,10 に設定した.

$$P(t) = \begin{cases} Power(t) & (Power(t) - Power(t-1) >= P) \\ 0 & (otherwise) \end{cases} \quad (3.3.9)$$

$$M(t) = \begin{cases} DoM(t) & (DoM(t) - DoM(t-1) >= M) \\ 0 & (otherwise) \end{cases} \quad (3.3.10)$$

3.4　評価実験

3.4.1　実験設定

本実験では，提案手法を用いることで自動生成される動作の印象を評価する．従来の機械学習を用いた音声駆動システムは，発話時の人間の動きを復元する手法である．そのため，本実験では従来手法の理想的な出力である発話時の人間の動きをアンドロイドにマッピングしたものと比較することで，提案手法 (Proposed with restriction,Proposed without restriction) と従来手法 (Copy) との比較とした．また，発話と動きが同期していない場合は，印象の悪化を招くことを確認するために，発話時の人間の動きに 1 秒の遅延をかけた条件 (NoSync) とも比較した．

本実験では，他者とのインタラクションを伴わない状況での発声動作を評価するために，アンドロイドが一方的に発話する様子を実験参加者が評価した．比較に用いる動作データ，音声データは節 3.3.4 で収録したものを用いた．発話の仕方に個性がある可能性があるため，本実験では節 3.2 と同じ 2 人の女性のデータを実験に用いた．図 3.6(b) と図 3.6(a) の Original データを比較すると話者 M のほうが動きの振幅が大きく，頻繁に

図 3.7 ビデオ刺激

動く傾向にある．この異なる個性を持つ人間の動作データと提案手法を比較することで，提案手法による生成動作の人間らしさの一般性を検証する．

Copy，NoSync 条件のものは動作データを用いて，Proposed(Proposed with restriction,Proposed without restriction) 条件は収録した音声データを用いて動作生成し，その様子をビデオ撮影し評価実験に用いた（図 3.7）．実験参加者は 4 条件を 2 人分の合計 8 条件すべて評価した．1 人目の動作パターン 4 条件を評価した後，2 人目の動作パターン 4 条件を評価した．各動作パターン 4 条件の順序はカウンタバランスをとった．使用する人間のデータの順序は印象形成に影響ないと考え，本実験では固定した．実験参加者にはビデオを繰り返し見ることを許可した．

3.4.2　評価指標

評価指標には，生成される動作が不自然にならないかを評価するために，「自然さ:首や腰の動きが自然であった」を 7 段階評価させた．さらに，従来知見より動作の自然さがエージェントに対する親密度を向上することがわかっているため [56]，対話に対する印象として，「対話意欲:このアンドロイドと話してみたいと感じた」「対話感:実際の人間が自分に向かって話をしているように感じましたか？」の 7 段階評価を用いた．

3.4.3　実験結果

実験参加者は 15 人（男性 12 人，女性 3 人，平均 21.5 歳，標準偏差 1.6 歳）であった．解析には人間データ要因（Human）・動作生成要因（Motion）の 2 要因被験者内分散分析を行った．対話意欲，対話感の項目では有意差は認められなかった．自然さの項目では，Motion 要因で主効果に有意傾向が認められた ($(F(3, 42) = 2.33) < 0.1$).

データを観察すると自動生成の 2 条件 (Proposed with restriction,Proposed without restriction) で，実験参加者によって評価にばらつきがある傾向が見られた．そこで，Proposed with restriction と Proposed without restriction のうち，実験参加者が期待する動作が個人ごとに異なると考えられるため，参加者にとってベストな生成手法を提案手法の評価値 (Proposed) として採用することにした．Motion 要因を 3 条件 (Copy, Proposed, NoSync) にし，さらに分散分析を行った．

対話意欲の平均値を図 3.8 に示す．Motion 要因で主効果に有意差が認められた ($(F(2, 28) = 4.90) < 0.05$)．Holm の多重比較を行ったところ，Proposed で Copy，NoSync より平均値が有意に高いことがわかった ($p < 0.05$).

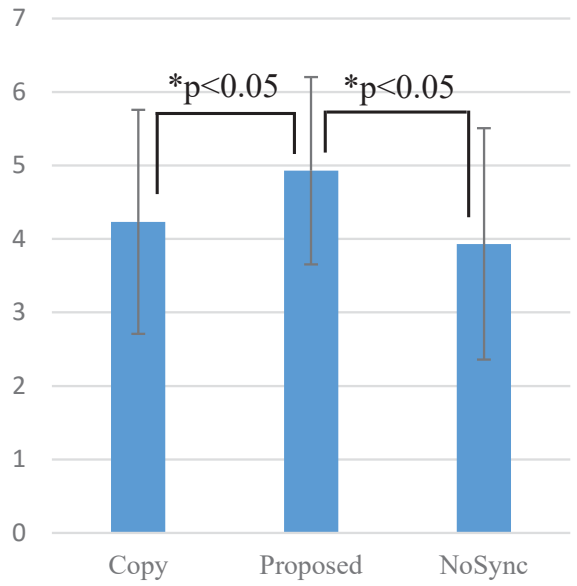


図 3.8 対話意欲

対話感の平均値を図 3.9 に示す．Motion 要因で主効果に有意差が認められた

$((F(2,28) = 6.51) < 0.01)$．Holm の多重比較を行ったところ，Proposed で NoSync より平均値が高い傾向にあることがわかった $(p < 0.1)$．

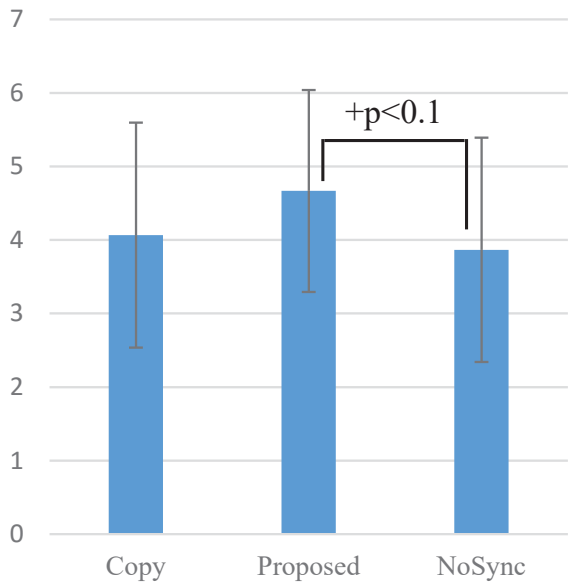


図 3.9 対話感

自然さの平均値を図 3.10 に示す．Motion 要因で主効果に有意差が認められた $((F(2,28) = 6.51) < 0.01)$．Holm の多重比較を行ったとこと，Proposed で Copy$(p < 0.05)$，NoSync$(p < 0.01)$ より平均値が有意に高いことがわかった．また，自然さでは交互作用も認められた $((F(2,28) = 5.89) < 0.01)$．図 3.11 に各条件ごとの自然さの平均値を示す．NoSync 条件にて，話者 H が話者 M より有意に高いことがわかった $((F(1,14) = 12.64) < 0.01)$．話者 M では，有意に Motion 要因の単純効果が認められた $((F(2,28) = 14.40) < 0.01)$．Holm の多重比較を行ったところ，Proposed で Copy$(p < 0.05)$，NoSync$(p < 0.01)$ より平均値が有意に高いことがわかった．また，Copy で NoSync 有意に高いことも認められた $(p < 0.05)$．

3.5　考察

節 3.4 の実験より提案手法を用いることで，実際の人間の動きをアンドロイドで表現するよりも自然にかつ対話意欲が出るような動作を生成できることがわかった．また，提案手法を用いると話者に依存せずに人間らしい動作が生成できることもわかった．

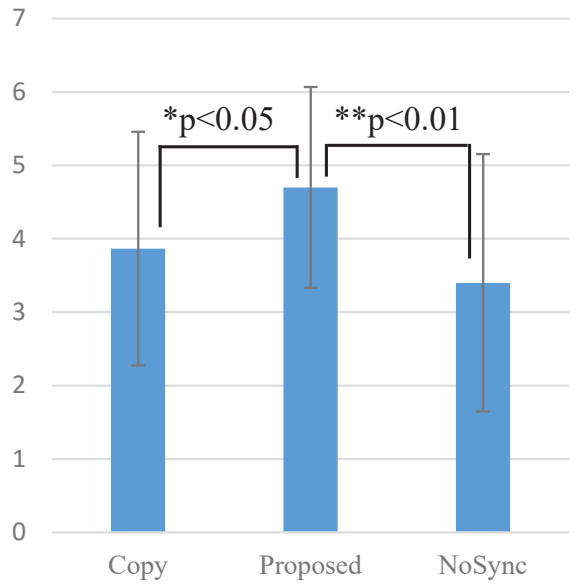


図 3.10 動きの自然さ

実際の人間はロボットのように首にも複数の関節があり，また表情などが常に微小に動作している．しかし，実際のロボットにはハードウェアの制約があるため，人間の動きを完全に再現することは困難である．そのため，今回の実験のように人間の動きをそのままロボットに写像した条件が低く評価されたと考えられる．一方，提案手法では，音声特徴と相関のある典型的な動きを明確に再現することで，ロボットを用いることで欠如する微細な動きによる人間らしさに対する印象を埋めていると考えられ，提案モデルが発声に伴う動きの人間らしさの特徴を捉えたモデルであると考えられる．従来の CG 分野での動作生成手法の多くが実際の人間の動きを再現することを試みている．それら手法では今回の人間の動きを写像する条件と同様に不自然さが生じると考えられ，実際のロボットに適応することは不適切である．また，微細な動きを顕著にするために，写像する関節角の倍率を高める手法などが考えられるが，常に人間が行っている微細な動きをロボットが行うとロボットの体全体が共振し不自然な印象を与えかねない．そのため，本研究のようにどのような動きモデルがロボットで表現でき，人間らしい印象を与えるかという知見はヒューマンロボットインタラクションを向上させるために重要である．

実験結果より，ほとんどの評価項目で Copy と NoSync で平均値に差があるが，統計的有意差が見られなかった．しかし，図 3.11 に示すように，NoSync 条件では人によっては顕著に自然さが低下することがわかる．これは話者 M は音声と動きに相関が強い傾向

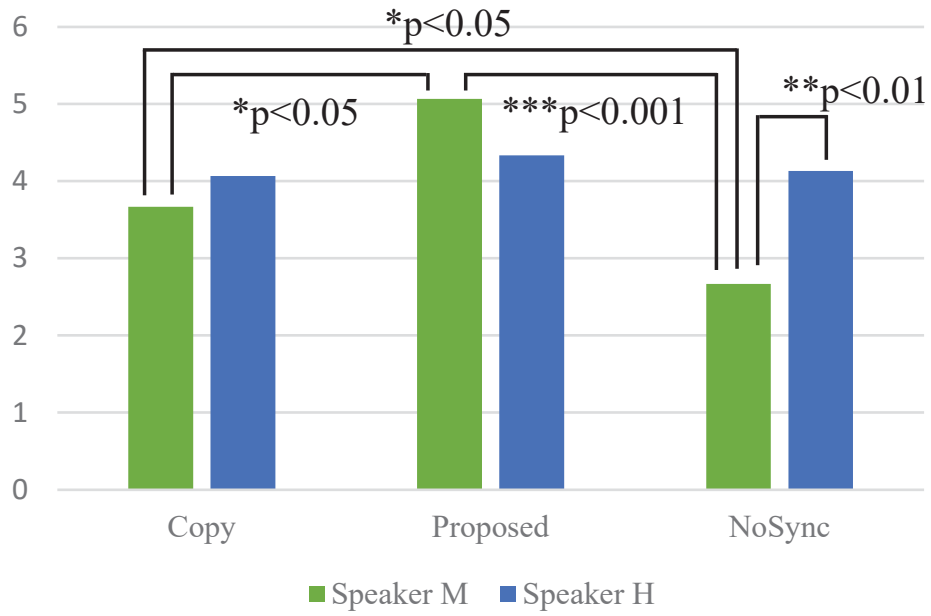


図 3.11 話者ごとの動きの自然さ

がある一方で，発声と動きがずれたとしても，頭部の動きに発声以外の要因を帰属したため，話者 H では NoSync 条件の印象が悪化しなかったと考えられる．つまり，発声に伴う頭部の動きが顕著に現れている場合は，発声との強い同期性が必要になる．

図 3.6(b) と図 3.6(a) の Original データを比較すると話者 M のほうが動きの振幅が大きな傾向にある．つまり話者 M のほうが，音声と相関が高くかつ大きく動く人であることがわかる．しかし，統計的有意差は認められなかったが，自然さの項目にて話者 M のほうが話者 H よりも平均値が小さいことがわかる．人間の体は全身で共振しないような制御が自動で行われているが，首の動き，腰の動きだけなど部分的な動きをその人以外の物理特性（異なる関節の動特性，身体部位の重さなど）を持つロボットに写像したため，単純に大きな動きを発話に合わせ行うと不自然になってしまったと考えられる．このように実際のロボットで動作生成する場合は，ロボット自身の動特性を考慮した，全身の関節動作の間の協調 (coordination) も重要になる．提案モデルでは，発声を起因とした関節間の動きの協調の特徴が再現されているので，実際の人間とは異なる動きでも自然と判断されたと考えられる．すなわち，自然さ動きには，身体的な拘束に基づく協調も重要と考えられる．

今回実験では実験参加者ごとに with restriction と without restriction に好みがばらつき

が見られた．実験参加者の自由記述を見ると，「動きが多かった」「動きが大きかった」から良いと評価する人や，逆に「動きが無駄に多い」から不自然と評価する人が見られた．カメレオン効果 [97] で知られるように人間は自分と同じような動きを好む傾向にある．そのため，実験参加者は自分の動き方に近い方を選ぶ傾向にあると考えられ，対話者の個性と適切な動き方について追究することが将来的に重要である．

3.6　展望

節 3.3 で提案した動作生成システムは 2 次遅れ系フィルターを用いて動作生成を行っている．2 次遅れ系であるバネ–ダンパ系を用いた筋肉のモデル化が行われていることから，緊張度合・感情状態によって変化する筋肉の硬さに応じた動きの変調も，筋肉の動特性に類似した 2 次遅れ系のフィルターでは実現できると考えられる．第 5 章にて, 発話時の緊張・感情状態にあった動作を生成することができるかの検証や直感的に動作パラメータを決定できるかのユーザビリティの面から提案手法を評価する．

提案手法では韻律特徴と頭部動作の基本的ルールに従い動作生成を行っている. しかし，実際の人間の動きは複雑であり，偶発的な要素が加わることで生物らしさが増す [98] ことが報告されていることから，ランダムに要素を動作生成に加えることでより生物らしさを強めることが出来ると考えられる．

3.7　まとめ

本章では，アンドロイドの発話音声に合わせリアルタイムで体幹動作を生成するシステムを提案した．韻律特徴と動作の相関を顕著にすることで，実際の人間の動きよりも自然な動作と評価されることがわかった．第 5 章では提案システムをアンドロイドの感情状態に適した動作に調節する手法を提案する．

第 4 章

発話の意味と一致するジェスチャー

4.1　まえがき

遠隔操作型アンドロイドでは，操作者の動きをアンドロイドで再現するが，操作インタフェースが介在することで，操作者の意味的動作が普段の対面時よりも減少する．それらの減少した動作をアンドロイドに付加することで円滑な遠隔対話が期待される．そこで，第 4 章では，発話の意味と強く相関する意味的動作を発話に合わせて表出することが，操作されたアンドロイドの人間らしさを向上させ，円滑な対話が実現できるという考えの下，操作者の音声から意味的動作を自動生成するシステムを構築する．動作としては，同意を表す頷きや，困惑を表す首傾げなど，コミュニケーションに重要な頭部動作を対象とする．

様々な意図表現動作を生成するためには，発話の意味情報が必要と考えられる．頭部動作と談話機能の相関に基づく Liu et al. の頭部動作モデル [65] は，発話の意味を表す談話機能を用いるため，日本語においてもいくつかの意図を表現する頭部動作を生成できる．Liu et al. の評価実験では，談話機能の識別が手動で行われており，遠隔操作ロボットに実装できるシステムは構築されておらず，インタラクションへの影響が検証されていない．談話機能の認識は音声認識の分野でも扱われていない課題であるが，感嘆詞を含む談話機能であれば既存の認識手法でも可能である．第 4 章では，操作者の発話中の

境くりま, 石井カルロス寿憲, 港隆史, 石黒浩, ”音声に対応する頭部動作のオンライン生成システムと遠隔操作における効果”, 電子情報通信学会和文論文誌 *A*, vol. J99-A, no. 1, pp. 14-24, January, 2016.
copyright©2016 IEICE

感嘆詞を含む談話機能を自動で推定し，[65] のモデルに基づいて頭部動作（頷き，首傾げ，首上げ）を生成し，遠隔操作ロボットに付加するシステムを構築し，このシステムが遠隔対話において有用であることを評価実験により検証する．そして，自動動作生成システムの評価を通して，感嘆詞に伴う発話と動作の同期が人間らしい動きと知覚させるために重要であることを検証する．

4.2　頭部動作システムの構築

4.2.1　談話機能と頭部動作の関係性に関する知見

Ishi et al. はマルチモーダル対話音声データベース (7 名（男性 4 名，女性 3 名）の話者による 20 個の自由会話（1 対話当たり 10 分から 15 分程度）) を用いて，頭部動作と談話機能の関係性を解析した [64),65)]．データベースは，各話者の発話がフレーズ単位で区切られている．フレーズ単位は，アクセント核を最大 1 つ有する音調単位となるアクセント句に相当し，文構造の面では複数の文節を含む場合がある．データベースには，それぞれのフレーズに対し，Ishi et al. が提案した下記の談話機能タグ [99),100)] が付与されている．

- k(keep): ポーズないしは，はっきりしたピッチのリセットが伴う強い句境界によって，話者が発話権を保持する．
- k2(keep): 発話の中にある弱い句境界によって発話権を保持する．
- k3(keep): 話者が発話末の音節を伸ばすことで，考えていることや発話の途中であることを表現し発話権を保持する．
- f(filler): 「えっとー」「あのー」など，考え中であることを表現する．
- f2(conjunctions): 「じゃ」などの接続詞で，文末を伸ばしていない短いフィラーとしてとらえられる．
- g(give): 当話者の発話が終了し，発話権を対話相手へ譲渡する．
- q(question): 対話相手に確認するなど応答を求め，発話権を譲渡する．
- bc(backchannels): 「うん」「はい」などの相槌を表現する．
- su(admiration/surprise/unexpectedness): 対話相手へのレスポンスとして「えー！」「うそ！」「へー」など，驚きや感心などを表現する．
- dn(denial, negation): 「いいえ」「ううん」などの否定を表現する．

談話機能と頭部動作の関係性の分析によると，頷き動作が対話の中で最も多く生起し，特に相槌 (bc) や強い句境界 (k,g,q) で多く見られることが報告されている [63]．また，発話者が考えていたり，次の発話の準備をしているなどの場合では，語尾を延ばすことが多い．それら弱い句境界 (f,k3) では，首傾げ動作が最も多く出現したことも報告されている．さらに，驚きや感心を表す感嘆詞 (su) においても，顔上げ動作や首傾げ動作が頻繁に見られる．

4.2.2　談話機能と発話の関係性に関する知見

日本語対話における談話機能と言語情報・韻律情報の関係性は [99),100)] において分析されている．相槌 (bc) は「うん」「ええ」「ああ」「はい」のような感嘆詞を下降調トーンで発話することが多く，驚きや感心 (su) は「ええ」「へえ」「うん」を上昇トーンで発話することが多いと報告されている．また，フィラー (f) は「ええ」「へえ」「ううん」などを，発話の保持 (k3) は「〜けど」「〜で」などの句末を長く平坦調に発話することが多い． このように，同じ言語情報である「ええ」であっても，韻律特徴であるイントネーションを下降調で発音すると同意を表し，平坦で発音すると不満や戸惑いを表し，上昇調で発音すると驚きの意味になる．これらの知見に基づき，本論文では，”bc”, ”su(admiration/surprise)”, ”f”, ”k3” の談話機能を自動的に推定する．

4.2.3　言語情報の抽出

4.2.2 節で説明したように，談話機能の推定には言語情報と韻律情報が必要となる．本論文では，オープンソースである大語彙連続音声認識エンジン Julius[101)] を用いて操作者の音声から言語情報を抽出する．Julius に付属する音響モデルは読み上げ音声を用いて作成されている．しかし，自然会話の「ああ」「ううん」「ええ」などの感嘆詞は，はっきり発音されないことが多いため，付属の音響モデルでは正しく認識することが困難である．感嘆詞の認識率を向上させるために，自然対話データベース [64)] の音声データから感嘆詞を抽出し音響モデルを作成した．音響モデルの学習には，4406 フレーズ（男性:1903，女性:2503）の音声データを用いた（「ああ」「ええ」「はい」「はあ」「へえ」「ほお」「うわあ」「わあ」「ううん」「いや」「いいえ」）．感嘆詞のモノフォン HMM（隠れマルコフモデル）の作成には HTK(`http://htk.eng.cam.ac.uk/`) を用いた． 韻律特徴は 12 MFCC（メル周波数ケプストラム），12 delta-MFCC，1 delta-power を用い，構築した音響モデルは

記述文法

$S : NO_B\ SENTENCE\ NO_E$
$SENTENCE : SENTENCE\ WORD$
$SENTENCE : SENTENCE\ FILLER$
$SENTENCE : SENTENCE\ NOISE$
$SENTENCE : WORD$
$SENTENCE : FILLER$

図 4.1 記述文法

1 感嘆詞ごとに 1 つのモデルを作成し，感嘆詞の長さに応じて HMM の状態数を 8~16 とした．

今回着目している感嘆詞は Julius に付属する新聞から作成した言語モデルには含まれないため，図 4.1 のような 1 文が感嘆詞とそれ以外の音声で構成される記述文法を用いた．単語辞書である WORD は今回作成した感嘆詞の音節で定義され，FILLER は五十音の音節すべてで定義した．音節は「か=k a」のような単純な音素の組み合わせで定義した．

4.2.4　韻律情報の抽出

4.2.2 節で説明したように，談話機能を推定するには韻律情報（音調）も必要となる．そのため，音声の基本周波数 (F0) を用いて音調の識別を行った．

まず，F0 の値の抽出には，32 ms のフレーム幅で 10 ms 毎に LPC(Linear Predictive Coding) 逆フィルタによる残差波形の自己相関関数の最大ピークに基づいた処理を行う．さらに，人間のイントネーションの知覚特性と一致するよう，F0 の値を対数スケールに変換した．

$$F0[semitone] = 12\log_2(F0[Hz]) \tag{4.2.1}$$

次に，感嘆詞内で F0 の変化量を表す $\Delta F0$(人間の音調の知覚に基づくパラメータ [102]) を抽出した．$\Delta F0$ は感嘆詞末の F0($\hat{F}0$) と前半部の F0 平均値 ($\overline{F0}$) との差分を用いて計算する（式 4.2.2）．$\hat{F}0$ は，感嘆詞後半部の F0 の近似直線の感嘆詞末での値である．そし

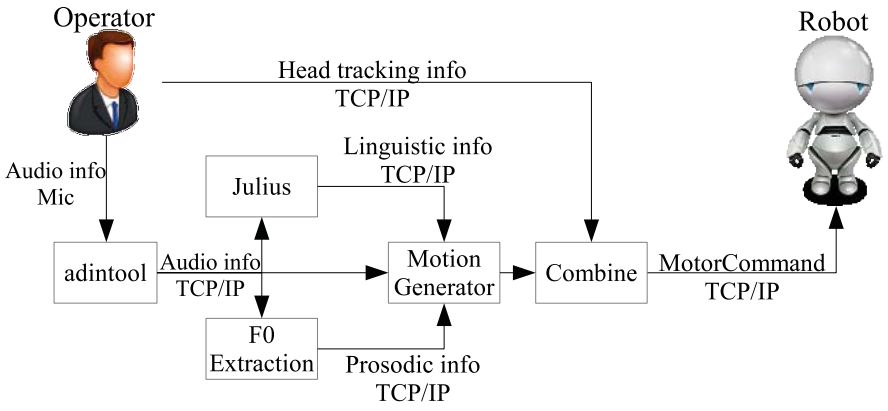


図 4.2 オンライン頭部動作生成システムの概要

て，感嘆詞の音調は式 4.2.3 に応じて，上昇調，下降調，平坦調に分類した．

$$\Delta F0 = \hat{F0} - \overline{F0} \tag{4.2.2}$$

$$tone = \begin{cases} rising\ (Rs) & (\Delta F0 > 1) \\ falling\ (Fa) & (\Delta F0 < -2) \\ flat\ (Ft) & (otherwise) \end{cases} \tag{4.2.3}$$

4.2.5　リアルタイム音声駆動頭部動作生成システム

図 4.2 に実装したシステムの概要を示す．adintool（Julius に付属）はマイクロフォンから操作者の音声信号を取得し，音声のパワーと零交差に基づき音声区間のセグメンテーションを行う．音声情報は言語情報を取得するために Julius に送られ，また韻律情報を取得するために F0 抽出部へ送られる．リアルタイムで処理するために，Julius は 100ms 毎に漸次認識結果を出力する（漸次認識結果に音素アライメントを出力させるよう Julius のソースコードを改良した）．F0 値は 10 ms 毎に計算される．動作生成部では，Julius からの音素アライメント情報に基づき，F0 情報を用いてキーワード区間の音調を識別する．抽出した言語情報と韻律情報に基づきロボットの頭部動作を生成し，ロボットにモータコマンドを送信する．すべてのモジュール間のデータ通信は TCP/IP を用いた．

図 4.3 に動作生成部のシステムフローを示す．Julius の認識結果は 100 ms 毎に出力されるが，出力毎に談話機能を推定すると，同一談話機能に対して重複して動作生成する可能性がある（「えー」のように平坦部が続くフィラーなど）．そのため漸次認識結果の感嘆

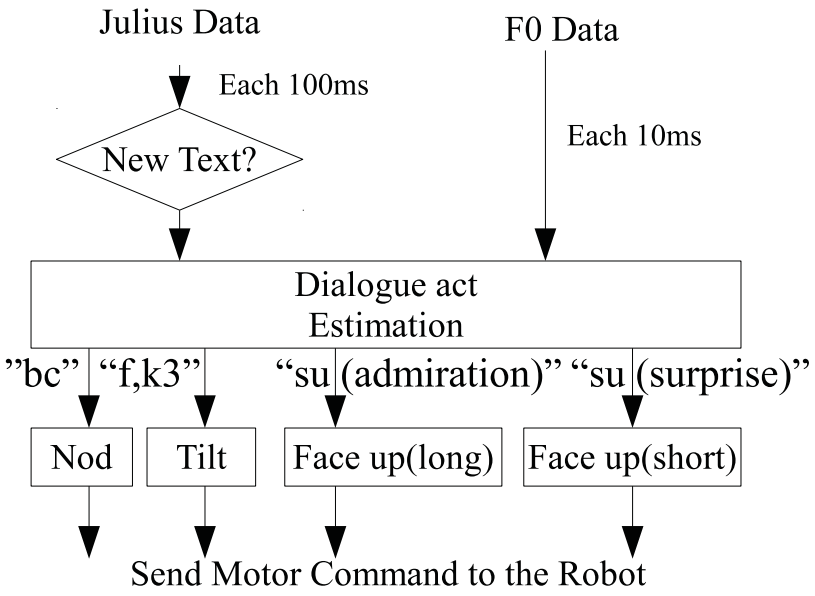


図 4.3 動作生成モジュールのシステムフロー

詞・音素が，すでに処理されたものの一部であるかどうかを確認するルールを設けている．抽出した言語情報と韻律情報に基づき頭部動作を生成し，それをロボットで実現するためのモータコマンドを生成する．音声処理と並列して，本システムは操作者の頭部動作を運動計測装置を用いて計測しており，それをロボットで実現するためのモータコマンドも生成する．動作統合部では，それら 2 つのモータコマンドを統合してロボットに送信する． ここでは単純に，操作者の動きを写像した動作に音声情報より自動的に生成された動作を加算し重ね合わせる．

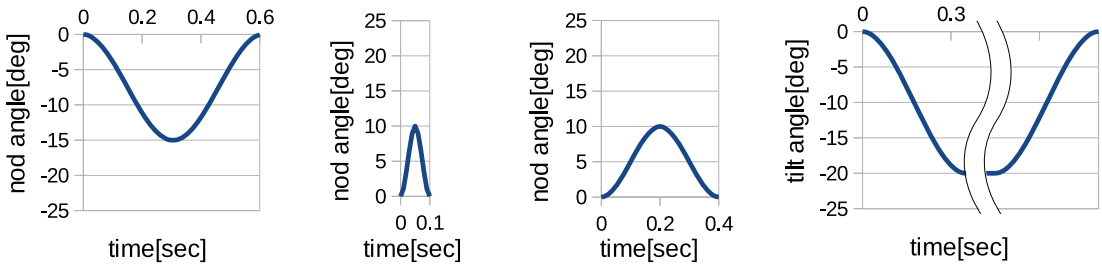
以下に各談話機能の推定と動作生成のルールを説明する．“bc” の推定は,「ああ」「あー」「ええ」「はい」「はあ」「へえ」「ほお」「ううん」「うん」のいずれかが認識され，その発話区間の音調が下降調である場合とする．“su(admiration)”，“su(surprise)” の推定は，「ええ」「へえ」「ほお」のいずれかが認識され，その発話区間の音調が上昇調である場合とする．発話区間が短い場合が “su(surprise)” であり，発話区間が長い場合が “su(admiration)” とする．“f” と “k3” の推定では，一般的な会話での母音区間の長さは 200~300 ms であるため，1 音節の閾値を 350 ms とし，1 音節が閾値以上長く平坦調である場合を “f”，“k3” した．表 4.1 に生成動作，談話機能，言語情報の間の対応関係を示す．

図 4.4 に生成する各動作の首の角度の時系列を示す．談話機能が推定されるとすぐに

表 4.1 音声情報と頭部動作のマッピング

		音声情報	
動作	談話機能	言語情報	韻律情報
頷き	bc	“ああ”, “ええ”, “はい”, “はぁ”, “へぇ”, “ほぉ”, “ううん”, “うん”	下降調
首傾げ	f, k3	1 音素	長い平坦調
顔上げ	su(surprise)	“ええ”, “へぇ”, “ほぉ”	短い上昇調
	su(admiration)	“ええ”, “へぇ”, “ほぉ”	長い上昇調

動作を生成し始める．動作の継続時間は従来知見 [64),65)] を参考に決定した．人間の頷き動作には，わずかに頭部を上げてから下げる動きが観測されている [63)]．しかし，動作生成の遅延をできる限り小さくするため，本実験では微少な首上げ動作を省略し，首下げ動作のみを実装した (図 4.4(a))．動作の振幅は，動きが明確に認識でき，かつ動作生成の遅延を小さくするためできるだけ小さい量とした．従来研究 [65)] 同様に，首傾げ状態は発話が終わるまで維持される．



(a) 頷き　(b) 顔上げ（驚き）　(c) 顔上げ（感心）　(d) 首傾げ

図 4.4 生成される頭部動作の時系列グラフ

4.3　提案システムの評価実験

4.3.1　実験目的

提案システムは，減少した操作者の頭部動作を補うことができるが，付加する動作そのものは操作者本人とは異なる動作である．そのため本実験では，本システムにおけるロボットの動作が自然であるかどうか確認する．さらに，自然な動作を実現したことの効果として，対話が円滑になるかどうかを評価する．

4.3.2　実験設定

本実験では遠隔操作ロボット「テレノイド」を用いた（図 4.5）[103]．テレノイドは首に 3 自由度（ピッチ軸，ロール軸，ヨー軸）と口に 1 自由度（上下開閉）のアクチュエータを持つ．テレノイドの外見は操作者に対する印象形成を乱さないような外見になっているため [104]，テレノイドを用いることで操作者の音声とロボットの外見との不適合による印象への影響を軽減することができる．また，テレノイドの口の動きは発話者音声に基づいて口唇動作を制御する手法 [41] を用いた．

本実験では，被験者に操作者（実験協力者）が操作するテレノイドと対話させ， 対話後に操作者の様子をどれほど把握したかを評価させることで，遠隔対話が円滑になったかどうかを評価した． 操作者の個性がもたらす印象のばらつきを統制するために，操作者は 1 人（女性）とした．

本実験では，操作者の頭部動作をそのままテレノイドで再現した “Copy” 条件と，“Copy” 条件に自動生成した動作を付加した “Copy+Auto” 条件を比較した．頭部動作は操作者の頭頂に取り付けた Inertial Measurement Unit (IMU) で計測した．“Copy+Auto” 条件では単純に操作者の動作に生成動作を加算した．

被験者が話し続けられるようにするために，事前に被験者に「好きなスポーツ」,「好きな映画」,「好きな本」の 3 つから興味のある対話のテーマを 2 つ選ばせた．また，本実験の評価のためには，被験者がテレノイドを見ずに話すような状況を避ける必要がある．そこで被験者が話を率先して話すように，被験者が話し始めるように教示した．この教示により，被験者はテレノイドから伝わる非言語情報から自分の話題に対する操作者の反応を把握しようとしてテレノイドに注意を向けることが期待される．被験者には 2 条

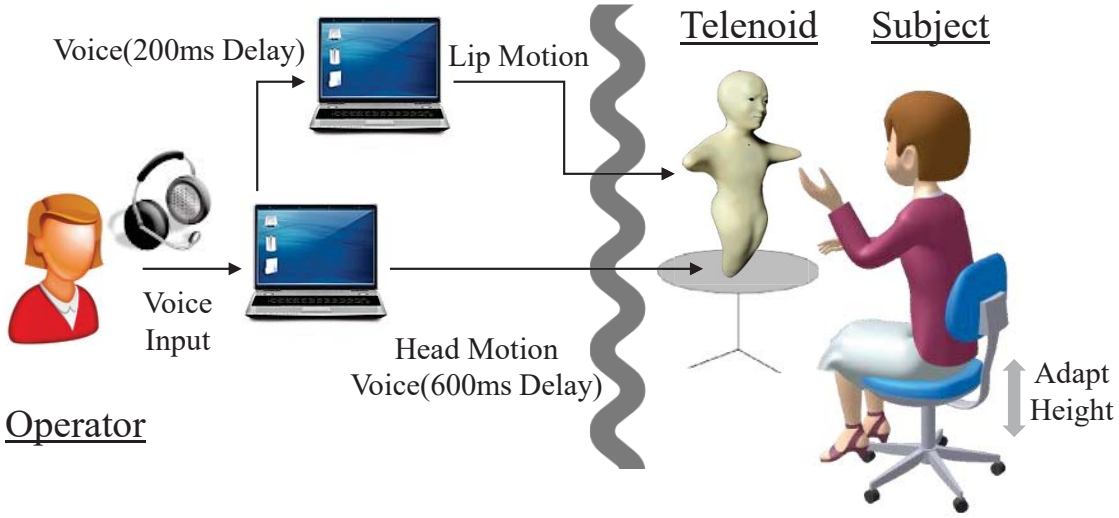


図 4.5 実験システム

件とも評価させ，被験者内比較を行った．条件の順序は被験者間でカウンタバランスをとった．また，操作者の行動に偏りが生じないよう，操作者には自分がどちらの条件でテレノイドを操作しているか分からないようにした．

4.3.3　実験システム

図 4.5 に実験システムを示す．提案システムの動作生成には音声認識による遅延が生じる．そのため，頭部動作と音声を同期させるためにテレノイド胸部のスピーカから操作者の音声を出力する際に 600 ms の遅延を人工的に設けた．操作者の反応が対話者に 600 ms 遅れて伝わるため，対話者の反応も 600 ms 程度遅れて操作者に伝わることになる．この遅れがあっても操作者が自然に対話可能かどうか確認するため，事前に操作者に体験させたところ，自然に対話できることが確認できた．また，口唇動作の生成に要する時間と頭部動作生成に要する時間が異なるため，口唇動作と頭部動作を同期させるために，口唇動作部への音声入力に 200 ms の遅延をかけた．ロボットの見え方を統制するために，被験者には椅子に座った状態でテレノイドの視線と自分の視線の高さが同じになるように椅子の高さを調節させた（図 4.5）．

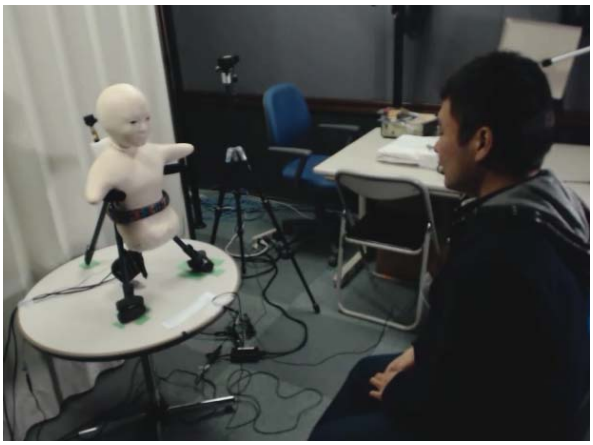


図 4.6 実験参加者と遠隔操作型アンドロイド

4.3.4　実験手順と評価指標

実験手順を以下に示す．

1. ロボットの説明と遠隔操作の実演：ロボットが自律型ではなく，遠隔操作で動いていることを被験者に教示する．
2. 3 分間の自由対話：被験者にロボットを通した遠隔対話に慣れさせる．
3. 1 回目の対話（3~5 分間）
4. 2 回目の対話（3~5 分間）
5. アンケート記入

本研究では，ヒューマノイドロボットを用いてビデオチャットなどでは伝わらない対話相手の物理的な存在感やそれに伴う動きを提示することで，対話相手の様子を把握しやすくし，遠隔対話の円滑化を図る．そこで提案システムにより操作者の様子を把握しやすくなるかを評価することで，遠隔対話の円滑化度合いを測る．本実験では，操作者にはインタラクションを一定時間以上続けるために，なるべく相手の話を盛り上げるよう指示した．つまり，操作者は実験参加者の話に興味を持って盛り上げるように話をするため，実験参加者にとっては，操作者が話を盛り上げるために実験参加者の話に集中しているように見えることが，対話の円滑さに関係すると考えた．そこで，円滑さの指標の 1 つとして，対話への集中度合いを選んだ．さらに「どちらのロボットのほうが話をちゃ

んと聞いていたか」「どちらのロボットのほうが話に興味を持っていたか」といった対話相手の様子についても評価させ，複合的に円滑さを評価させた．また，相手と対面している感じがあると，手掛かりが多く入手でき円滑な対話に近づくと考えられるため，「どちらのロボットのほうが相手と対面している感じが強かったか」についても評価させた．提案システムでは操作者の動きとは異なる動作を生成するため，付加した動作が不自然と評価される可能性があるため，動作の自然さも評価させた． 被験者は 1 回目と 2 回目の動作のうちどちらが対話に適しているかを 7 段階（1: 1 回目の対話の方，4: どちらとも言えない，7: 2 回目の対話の方）で評価した．

4.3.5　実験結果

実験の被験者は 16 人（男 : 8 人，女 : 8 人，平均年齢 : 21.4，標準偏差 : 0.36）であった．

提案システムにより生成された動作の付加による 相手の様子の把握しやすさ の変化を検証するため，アンケート記入後に行った被験者へのインタビューにおいて，評価の違いをテレノイドの動作以外に理由づけた被験者を解析から外した．3 人の被験者が対話中ロボットを見なかったと答え，また 2 人の被験者が一方の アンケート結果の理由を， ロボットの表情が一方の条件だけよかったためと回答した．今回使用したテレノイドは表情を変化させる自由度はないため，動作の違いではなく音声に含まれる 感情から，対話相手の様子 を評価したと考えられる．これら 5 人の被験者は動作の付加効果を評価していない可能性があるため解析から外し，11 人（男 : 5 人，女 : 6 人，平均年齢 : 21.3，標準偏差 : 0.47）のデータを用いて解析を行った．残りの被験者は「動作が大きかった」「会話と動作がリンクしていた」など動作の違いが理由であるとインタビューに回答していた．

実験条件と対話の順序はカウンタバランスをとったため，アンケートの結果を “Copy+Auto” 条件が当てはまる場合が正に，逆の場合は負になるように変換した．このように変換した後，それぞれのアンケートのの平均点が 0 より大きいかどうかを検定した．その際，Shapiro-Wilk 検定により分布の正規性が認められた場合には t 検定，認められなかった場合には Wilcoxon の順位和検定を用いた．対話への集中度合いについて，“Copy+Auto” 条件が有意に高いことが認められた (平均 0.73, $p < 0.05$ ：Wilcoxon の順位和検定)．また，動作の自然さについても，“Copy+Auto” 条件が有意に高いことが認められた (平均 0.73, $t(10) = 2.03; p < 0.05$ ：t 検定)． 図 4.8 に示すその他項目においても，“Copy+Auto” 条件が有意に高い傾向が認められた (ちゃんと聞いていた度合; 平

均 0.55, $t(10) = 1.60; p < 0.1$ ：t 検定), (興味度合; 平均 0.73, $t(10) = 1.55; p < 0.1$ ：t 検定), (対面感; 平均 0.73, $p < 0.1$ ：Wilcoxon の順位和検定)．動作の自然さについては，両条件とも絶対値での評価も行わせており，“Copy+Auto” でのロボットの動作の自然さの平均値が 5.2，標準分散が 0.44 であり，“Copy” でのロボットの動作の自然さの平均値が 5.2，標準分散が 0.38 であった（評価は 7 段階評価で 7 がすごく当てはまる，1 が全く当てはまらない）．平均値が 4 以上であることから (“Copy” 条件, $t(10) = 3.13; p < 0.01$ ：t 検定)(“Copy+Auto” 条件, $p < 0.05$ ：Wilcoxon の順位和検定)，両条件とも被験者はロボットの動作自体が不自然であると感じてはいないことが分かった．

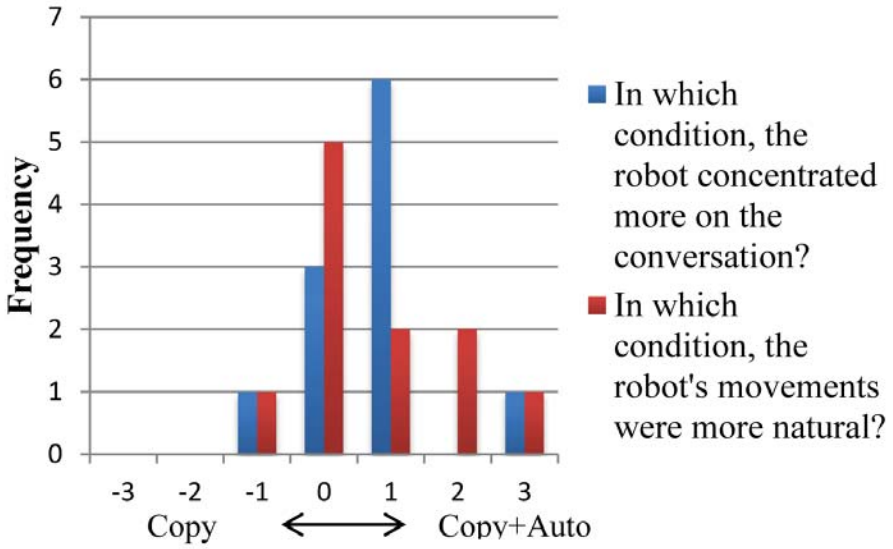


図 4.7 “Copy+Auto” と “Copy” の対比較ヒストグラム (有意差あり)

表 4.2 に生成すべき動作（期待される生成動作）に対するシステムが実際に生成した動作（実際の生成動作）の回数を示す．生成すべき動作は，記録した操作者の映像のアノテーション結果に基づいてカウントした．この表から，35% の動作が誤ったタイミングで動作を生成し，10% の動作が生成すべき時に生成されなかったことが分かる．

これら実験結果より，提案システムが操作者の頭部動作を自動的に補償することで，対話への集中度合いをより伝えることができ，対話の円滑化に貢献することが明らかになった．

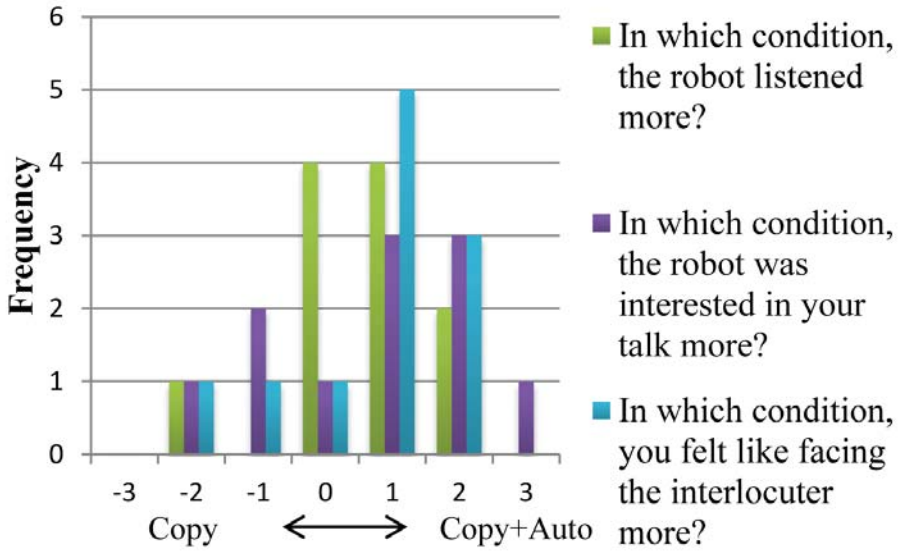


図 4.8 “Copy+Auto” と “Copy” の対比較ヒストグラム (有意傾向)

表 4.2 正解動作と生成動作の数

	期待される生成動作			
実際の生成動作	頷き	首傾げ	顔上げ	その他
頷き	**775**	2	77	127
首傾げ	48	**3**	13	13
顔上げ	132	0	**56**	38
動作生成なし	115	0	0	

4.4　考察

動作の自然さ，対話への集中度合いの結果を，図 4.7 に示すアンケート結果のヒストグラムからさらに解析する．動作の自然さについては，“Copy+Auto” 条件の方が自然と評価する被験者もいたが，ヒストグラムからわかるように評価値の最頻値は 0 であり，多くの被験者は条件間に差がないと評価している．一方で，対話への集中度合いについては，最頻値は +1 であり，“Copy+Auto” 条件の方がより対話に集中していると感じている被験者が多いことがわかる． すなわち，“Copy” 条件では，遠隔操作インタフェースにより

操作者の動作が減少している可能性があるにも関わらず，動作自体は不自然な印象を与えていない．ただし，実験前の予想通り，操作者の様子の把握しやすさは，頭部動作を自動的に補償しない場合より，劣っていることが分かる．また，動作の自然さの絶対値の評価結果より，ロボットの動作が操作者本人の動作と異なっている可能性があるにも関わらず，動作の自然さが失われていないこと，本人の動作と異なる動作で補っても操作者の状態を把握しやすくなることが分かった．

表 4.2 の結果が示すように，誤って生成された動作が 35% あるにも関わらず，集中している度合いなどの評価が向上していることは，例え誤った動作であっても動作が多い方が，対話相手に何かしらの意図を帰属させることができ，相手の様子を把握した気にさせる可能性を示していると考えられる．しかし，正しく操作者の様子を把握させるには，どこまで自動で動作生成しても，操作者の意図に反しないかを明らかにすることが重要である．従来研究では，操作者の考える理想の動きと操作者の実際の動きには差があることが報告されている [105]．つまり，操作者に，対話中に自分がどのような動作をするかを尋ねると，その答えの動作は，実際に操作者が対話中に行う動作よりも誇張された動作となる傾向があることが明らかになっている．そのため，自動的に生成した動作が，操作者の意図通りの動きかどうかを評価することが困難である（[105] によれば，例え対面対話時の操作者の動きをロボットに実装したとしても，それを見た操作者は，自分ならもっと動かすはず，と答える傾向がある）．しかし，上記で述べたように操作者の望む動きが現実の動きよりも誇張される傾向にあることから，提案システムにより生成された動作の方がより操作者の望む動きに近い可能性がある．

インタフェースによる操作者の動作の減少は，対話する両者ともロボットを用いることで防げると考えられる．しかし，現段階では両方ともロボットにした際に，どのようにロボットの全ての動作を生成するかが問題になる．提案システムは片方のみロボットの場合だけではなく，両方ともロボットの場合においても，音声のみから動作を生成できるため，どちらの対話スタイルにおいても有効な手法である．また，高齢者介護現場では，外出できない高齢者と遠隔操作ロボットを介して会話を行う傾聴ボランティアが行われている [106]．このような状況では，片側のみモニタを用いたインタフェースは，モニタで高齢者の様子を詳しく観察しながら対話することができ，高齢者支援という目的に適した対話スタイルになっていると考えられる．片側のみモニタを用いた対話スタイルが積極的に利用される需要があることを考えると，このスタイルでの問題点を解決することは意義があると言える．さらに，提案システムを使用すると操作者が動きを含め，対話に

対する負担を軽減しながらも，高齢者は対面時のような対話ができるため，より人手不足が進む高齢者介護の現場で役立つことが期待される．

4.5　展望

本実験では，相槌や感心などの 4 つの談話機能に同期する頭部動作ならば，遠隔操作アンドロイドに自動的に付加してもアンドロイドの動作の自然さを損なうことはないことが示されたが，どのような動作が人工的に生成した動作と置き換え可能であるか，すなわち，どこまで動作を自動的に生成しても自然さを損なわないのかは明らかになっていない．このことを明らかにすることは，遠隔操作型ロボットにより対話を支援するシステムを構築する上で重要な課題である．Liu et al.[65] は，本実験で用いた 4 つ以外の談話機能と頭部動作の対応もモデル化しており，それらを用いればさらに多くの動作を補償することが可能であるが，認識に形態素解析が必要な談話機能もあるため，認識処理のリアルタイム化が開発課題として残されている．また，何人かの被験者は，テレノイドの動きに視線の動きや感情が分かる動作（表情など）が欠如していることを指摘している．これらもコミュニケーションに重要な動作であるが，ビデオチャットのようなインタフェースを用いると欠如しやすい動作である．これらの動きを音声から補償するシステムの構築も重要な将来課題である．

また，提案システムは，談話機能に応じて頷き，首傾げ，顔上げの動作を自動生成できるものであったが，実際にこれらの動作が生起された回数を調べると，表 4.2 が示すように，首傾げの動作はほとんど生じていない．実験で設定した会話条件では，首傾げがほとんど生じない会話であったと考えられるため，首傾げ動作の効果については，会話を制限して，首傾げの頻度が高まる状態で，さらに評価を行うことが必要であると考えられる．

本実験における “Copy+Auto” 条件のロボット頭部動作は，音声情報より自動的に生成された動作と，操作者の動きを写像した動作の重ね合わせである．そのため，操作者が発話しながら頷く場合など，自動生成による頷きと写像による頷きが重ね合わされ，ロボットは深い頷きや，素早い二連続頷きを行うことがあった．そのような動作があったにも関わらず不自然と評価されなかった理由は，日常対話の中でもそのような頷きが現れるためだと考えられる．本システムでは単純な重ね合わせ手法を用いたが，様々な動作を自動で生成するためには，動作の自然さを損なわないように重ね合わせる手法を考案する必要がある．

本システムでは，生成した動作と音声を同期させるために音声に遅延をかけたことで，600 ms の伝達遅延が生じている．遠隔対話において対話に悪影響を及ぼす伝達遅延については，様々な実験結果が報告されている（300 ms[107]，600 ms[108]，800 ms[109]）．本実験は 600 ms の遅延がありながらも，遅延による対話しづらさを指摘する被験者はいなかった．遠隔操作型ロボットによる対話では，従来知見よりも大きな遅延が許容される可能性がある．本システムの自動化を発展させる上で，遅延の許容範囲を明らかにすることは重要な課題である．

4.6　まとめ

本章では操作者の音声に含まれる談話機能をリアルタイムで推定し，遠隔操作ロボットのいくつかの頭部動作を自動生成するシステムを提案した．このシステムを遠隔操作ロボットの動作生成に組み込み，操作者の欠如した頭部動作を補償することで 対話を円滑化できることを明らかにした． また，生成した動作は操作者の動作とは異なるものであるのにも関わらず，ロボットの動作の自然さを損なわないことが分かった．さらに，発話の意味に一致しない動作を自動的に付加しても評価が悪化しなかったことから，発話とその意味に一致する意味的動作を同期的に表出することが，人間らしさの確証バイアスを誘発していることを示していると考えられる．今後の課題として，どこまで動作を自動的に生成しても自然さを損なわないのか， また対話相手の様子を正確に把握できるのかを調査する必要がある． 本章の結果はモダリティ間の相関を利用して動作を自動生成する手法の実用性を示すものである．提案システムでは，音声と頭部動作の関係性を利用したが，音声から注視動作や表情を自動生成する手法も重要な研究課題である．さらに，音声と動作の関係性を利用すれば，ロボットと対面する者の音声や動作モダリティから自動でロボットの反応動作を生成することも考えられる．本システムを発展させることで，遠隔操作型ロボットのみならず，自律ロボットの動作生成への貢献も期待される．

第 5 章

感情的発話と一致した症状的非言語動作

5.1　まえがき

本章では，アンドロイドの発話音声から伝わる感情と一致した感情的な動きの生成に取り組む．感情レベルで発話と一致した動作を生成するシステムを構築するためには，まずは感情的な動き方がどのような特徴になっているかを解明し，次に感情的な喋り方と感情的な動きをどのようなタイミングで同期させると効果的に人らしい感情を表現できるかを明らかにし，自動動作生成システムの構築を行う．本論文では，第 1 段階である，感情的な動き特徴を明らかにする．

第 3 章にて提案した音声駆動頭部動作生成システム [70] では，発声に伴う無意識的な頭部動作の特性を人間の筋肉の硬さに相当するパラメータによって変更することができる．生理学的知見 [77)–79)] に基づけば，感情が筋肉の緊張・弛緩に影響を及ぼすことで，感情に応じて人の動き方が変化すると考えられる．そこで，提案システムのパラメータを適切に選ぶことで様々な感情を表現できると考えられる．第 5 章では，この頭部動作生成システムのパラメータ空間が感情空間と連続的にマッピングできることを実験によって示し，感情の連続的な変化に合わせて発話動作を連続的に変調することができるシステムを構築する．本論文では人間らしさの特徴を明らかにして動作生成のモデルを構築する

境くりま, 港隆史, 石井カルロス寿憲, 石黒浩, ”わずかな感情変化を表現可能なアンドロイド動作の生成モデルの提案”, 電子情報通信学会和文論文誌 *D*, vol. J100-D, no. 3, March, 2017. copyright©2017 IEICE

アプローチを取っているが，本章のシステムが有効に働くことを示すことで，本アプローチの意義を示す．

5.2　実験設定

5.2.1　実験目的

本実験では，第 3 章にて提案した音声駆動頭部動作生成システム [70] を用いて，人間の発話に伴う無意識的な発話動作（頭部の上下方向の動き）が感情に対応してどのように変化すべきかを明らかにすることを目指す．従来研究では，指定した感情を表す動きを人に演じさせ，その動きを解析する方法が主である．しかし，ある感情を表現する動きは，実際に自分で動く場合と，客観的にイメージする場合とで異なることが知られている [105]．本研究では，アンドロイドと対面する者がアンドロイドの動作から推測するアンドロイドの感情とその動作との関係を明らかにすることを目的としている．そのため本実験では，アンドロイドが指定の感情を表現していると被験者が感じるように，被験者自身が動作生成システムのパラメータを調節することで，感情状態と動作パラメータとのマッピングを明らかにする．

5.2.2　実験システム

式 3.3.1 のパラメータ J，D，K はそれぞれ頭部の重さ，筋肉の粘度，筋肉の硬さに相当する．先行研究から，感情状態と動きの大きさおよび速さが関係していると考えられる [53),54),74),76] ことから，これらの動きの特徴を調節しやすいように，式 3.3.1 の独立した 3 つのパラメータ J, D, K を式 5.2.1 のように変換する．ω_0 は固有角振動数 (natural angular frequency), ξ は減衰比 (damping ratio), ϕ は慣性の逆数 (the reciprocal of inertia) である．ω_0 は振動の収束の速さを表し，ξ は減衰の強さを表す（$\xi > 1$：過減衰，$\xi = 1$：臨界減衰，$\xi < 1$：減衰振動）．また，ϕ が大きいほど動きが大きくなる．

$$\begin{gathered}\ddot{\theta}_{base}(t) + 2\xi\omega_0\dot{\theta}_{base}(t) + \omega_0^2\theta_{base}(t) = \phi T(t)Dir(t) \\ J = \frac{1}{\phi},\ K = \frac{\omega_0^2}{\phi},\ D = \frac{2\xi\omega_0}{\phi}\end{gathered} \tag{5.2.1}$$

本実験では，周期的動作となる発話動作をターゲットとしているため，減衰比は $\xi \approx 1$ となるべきである．そこで減衰比を $\xi = 1$ と固定し，被験者には ω_0 と ϕ を調節させる．

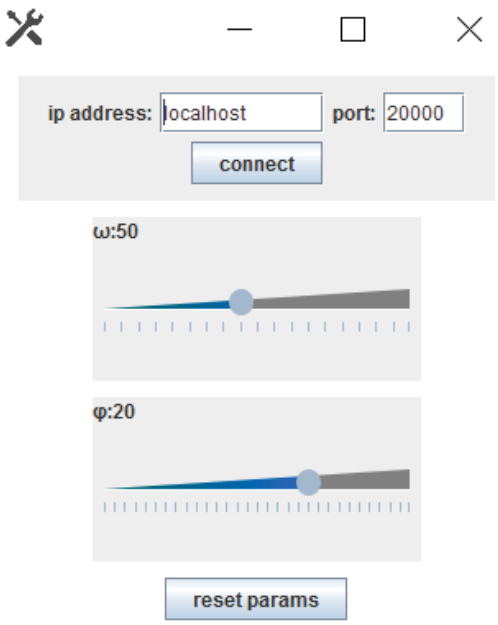


図 5.1 動作パラメータ操作インタフェース

操作インタフェースは図 5.1 のようになっており，ω_0 は 1 から 10 の範囲を 0.5 刻みで調節でき，ϕ は 10^0 から 10^2 の範囲を $10^{0.05}$ 刻みで調節できる．アンドロイドに発話させる音声をシステムに入力すると，アンドロイドの頭部スピーカからその音声が再生され，その音声に同期するようにアンドロイドの頭部が動作する．ただし，指令値の計算時間やアンドロイドとの通信時間等の時間遅れがあるため，音声を遅延 (333msec) させて再生することで，音声と動作を同期させる．被験者が ω_0 と ϕ を変化させると，アンドロイドの発話動作が変化する（発話中にパラメータを変化させると即時に発話動作が変化する）．被験者はその動きを見ながら，指定された感情を表現している動きになるパラメータを見つけ出す．

5.2.3　実験条件

感情状態として Russell の円環状モデル [71] の 4 象限からそれぞれ楽しい・退屈・リラックス・緊張の 4 つを選び，被験者にそれらの感情を感じる動作になるパラメータを見つけさせた．連続的な感情と動作のマッピングを明らかにすることが目的であるが，本実験では被験者には 4 つのシンボル（楽しい・退屈・リラックス・緊張）で感情を指定した．

ただし，各感情を感じる動きは被験者によってばらつきがあると考えられ，複数の被験者でパラメータを見つけさせることにより，退屈寄りの楽しい状態を表す動きや，緊張寄りのリラックス状態を表す動きなど，4 つの感情間の動きのパラメータも収集されると期待される．

本実験で使用する動作生成システムは，発話音声の韻律特徴を入力とするため，異なる音声（異なる発話）を入力にすると異なる発話動作を生成する．頭部動作と発話音声が被験者ごとや感情ごとに異なると，動作の変調させ方が統制されない可能性があるため，実験で使用する音声は，全ての被験者，全ての感情条件において同一の音声を用いた．音声は，女性の実験協力者が 1 分程度のニュース原稿を読み上げたものである．音声収録時には，女性実験協力者にニュートラルな感情で読むよう指示した．

予備実験から表情が変わらない状態では，感情に合わせた動きを調節することが困難であることがわかったため，アンドロイドには発話に伴う頭部動作に，各感情に合わせた表情と視線動作を加えた（パラメータを変更しても表情と視線動作は変化しない）．Ekman の知見 [110] に基づき，楽しい・リラックス条件では口角を上げ笑顔にし，視線を周期的に左右どちらかにそらす視線動作を加えた．また，退屈・緊張条件では目の開きを小さくし，視線を周期的に左下，右下どちらかにそらす視線動作を加えた．

5.2.4　実験手順

非言語情報から他者の個性や感情を判断する基準は，判断する人の個性に依存することが報告されている [111]．そこで被験者の性格診断を行うために，実験を始める前に NEO-FFI アンケートに回答させた．被験者は図 5.2 に示すようにアンドロイド ERICA（図 3.5）の前に座り，操作インタフェース（図 5.1）を用いて，指定された感情が感じられる動きになるようにパラメータを調節した．動作生成システムに音声が入力され動作がアンドロイドに反映されるまでに遅延が生じるため，スピーカーから出る音には 333msec の遅延を設けた．被験者には，満足する動作が得られるまで，繰り返し音声に伴う動作を見ながらパラメータを調節することを許可した．調節する感情の順序は，被験者間でカウンタバランスを取った．

また，動作生成システムのパラメータ調節の容易さを評価するために，客観的な調節の容易さの指標として調節に要する時間を計測した．さらに，被験者が調節した動作に指定した感情を感じているかどうかを評価するために，調節結果の満足度（思い通りの動作

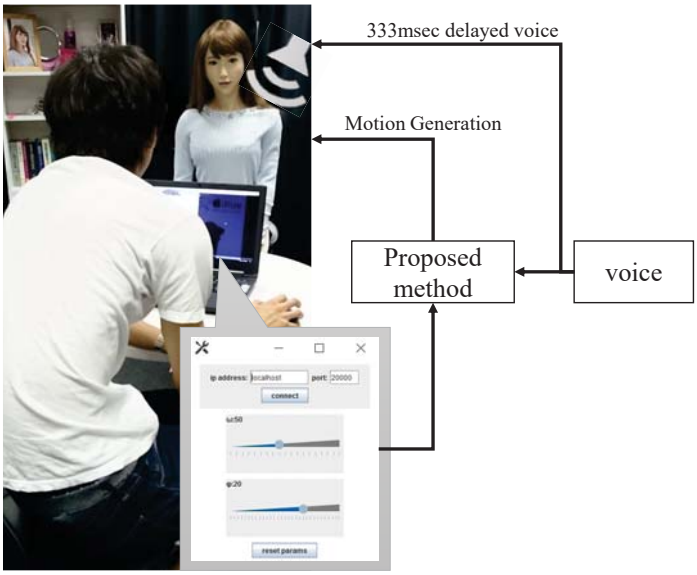


図 5.2 ブロックダイアグラム

を生成できたかどうか）を 7 段階で評価させた（1：不満足 ~7：満足）.

5.3　実験結果

実験には 12 人（男：6 人，女：6 人，平均年齢 20.4，標準偏差 1.0）が参加した．被験者は大学生対象の就職支援サイトで募集して集めたため，全員が大学生であるが在籍学部などの背景はそれぞれ異なる．

調節されたアンドロイドの動きの特徴を抽出するために，アンドロイドの頭頂部に取り付けた Inertial Measurement Unit (IMU) を用いて頭部のピッチ角（上下方向の角度）を計測した．感情表現には動きの大きさ，速さが関係していることから [53),54),74),76)]，頭部角度変化の大きさと速さを抽出した．速さは 10 msec ごとの角度変化量の絶対値として計算した．大きさは角度時系列の局所的な振幅（隣り合う極値の差の絶対値）として計算した (図 5.3).

調節されたパラメータを用いて 1 発話を行った際の動きの大きさと速さの中央値を，全被験者の全感情条件についてプロットしたものを図 5.4 に示す．このデータの分布特性を調べるために，データ分布を混合ガウス分布で近似した．混合ガウス分布は混合数を 1~3 の範囲でそれぞれ EM アルゴリズム [112)] を用いて推定し，赤池情報量基準 [113)] を

用いて最適な混合数を感情ごとに求めた．その結果，退屈，緊張では混合数 2 となり，リラックス，楽しいでは混合数 1 となった．図 5.4 に示される楕円は各ガウス分布を表し，半径はガウス分布の分散の平方根である．この結果から，データがある程度偏りを持って分布しており，被験者が異なっても，同じ感情を感じる動きとして，似た特徴の動きを見つけ出していることが分かる．さらに想定通り，被験者によっては，退屈寄りの楽しい状態を表す動きや，緊張寄りのリラックス状態を表す動きなどを見つけ出していることが分かる．

図 5.4 の 4 つの感情のデータを同一グラフにプロットし，線形近似すると高い相関が認められた（相関係数 0.84，図 5.5）．これらのことから，感情の変化はこの直線 θ_e 上の 1 次元空間上の動作変化に対応すると考えられる．さらに，図 5.4 のガウス分布を同一グラフに描画したものを図 5.6 に示す．図 5.6 の左下から右上にたどると，感情が緊張から退屈，リラックス，楽しいと遷移し再度，緊張に戻るというループ状に遷移することが分かる．これは，Russell の円環モデルにおいて反時計方向の感情遷移に一致し，図 5.7 に示すように Russell の円環モデルの反時計方向の感情は，θ_e 軸の 1 次元空間にマッピングできることがわかる．したがって，感情空間が連続的に変化するのに対し，その感情を表現する動き特徴も連続的に変化することがわかる．ただし，図 5.6 に示すように感情の境界部分では，同じ特徴の動きが複数の感情に感じられることがある．

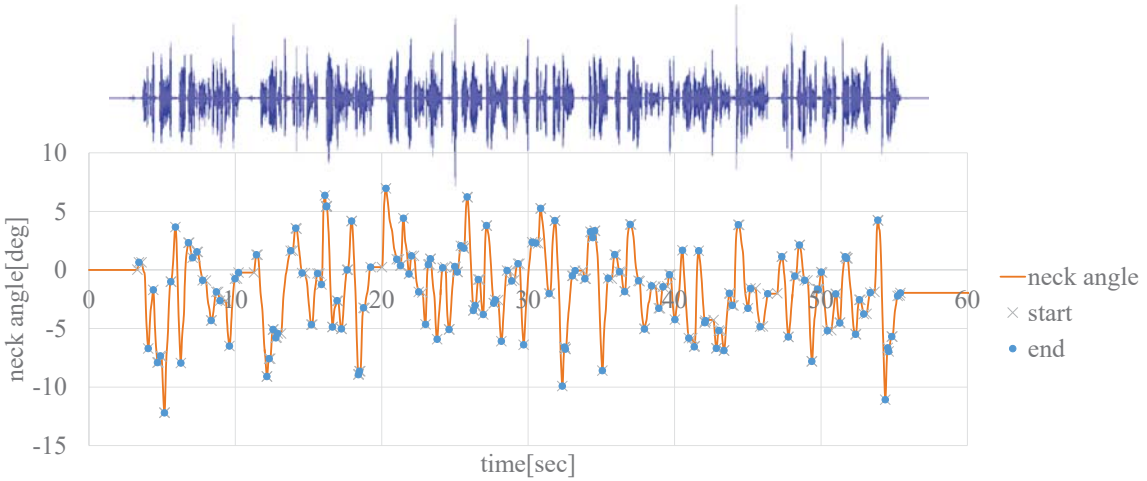
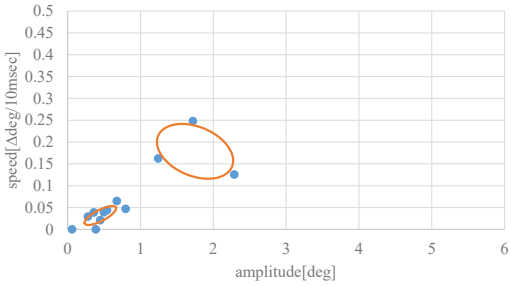
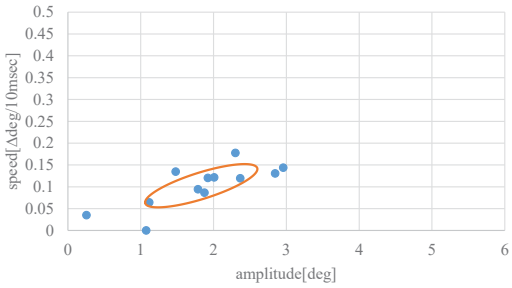


図 5.3　頭部動作の例と動きの大きさ特徴の抽出（start と end は隣り合う極値のペアを示している）．上のグラフは音声波形で，下のグラフは生成された頭部動作

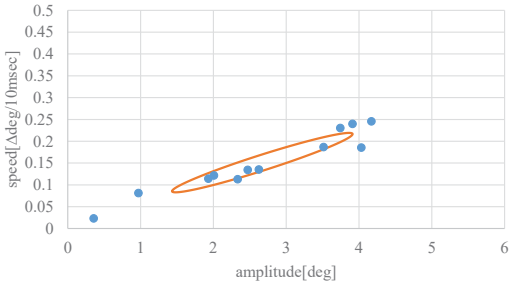
次に，動き特徴空間上の直線 θ_e を $\omega_0 - \phi$ パラメータ空間に写像する．ここでは，式 5.2.1 で生成される動作指令値が直線 θ_e 上に乗るようなパラメータを探索した．その結果を図 5.8 に示す．動き特徴が統計量（発話動作内での動きの大きさと速さの中央値）であるため，パラメータと動き特徴とは 1 対 1 に対応するとは限らない．そのため，動き特徴空間のある 1 点は，パラメータ空間上のある領域に写像される（すなわち，ある 1 点の感情を表現するパラメータが複数存在する）．ここで，すべての感情領域を通るように直線を定める（θ_e^p）．直線 θ_e^p 上に沿うようにパラメータを変化させると，下から順に，緊張，退屈，リラックス，楽しい，緊張を表現するように発話動作が変化する．緊張と退屈の中間状態や退屈とリラックスの中間状態のような感情も表現できる．このように，Russell の円環モデルの円周方向の感情の変化が，音声駆動頭部動作生成システムのパラメータ空間のある 1 次元上にマッピングできることが示された．したがって，円環モデル上の 1



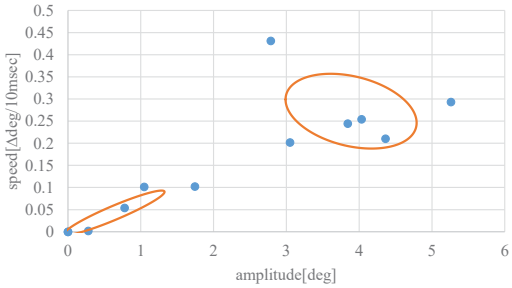
(a) 退屈



(b) リラックス



(c) 楽しい



(d) 緊張

図 5.4 感情と動き特徴の関係

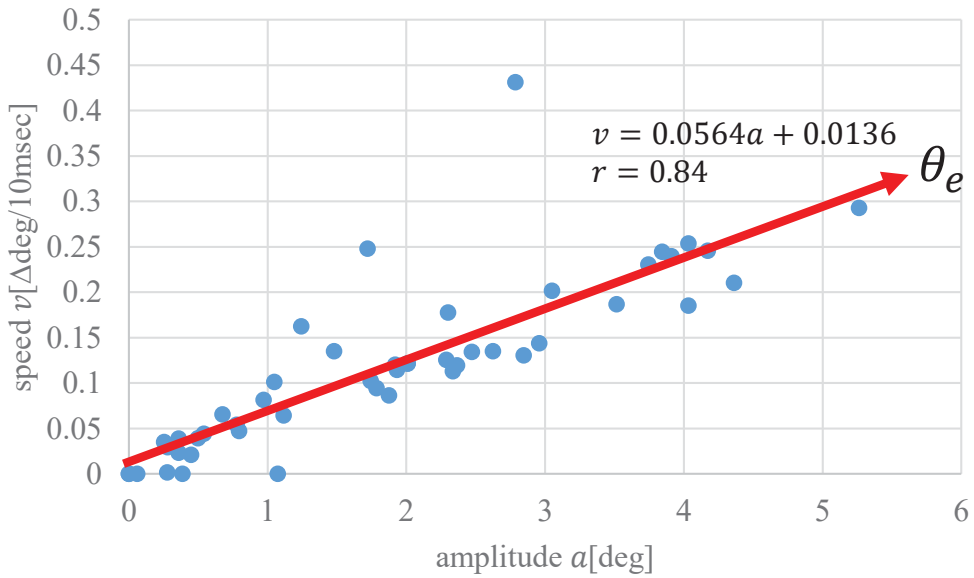


図 5.5 動き特徴の線形近似

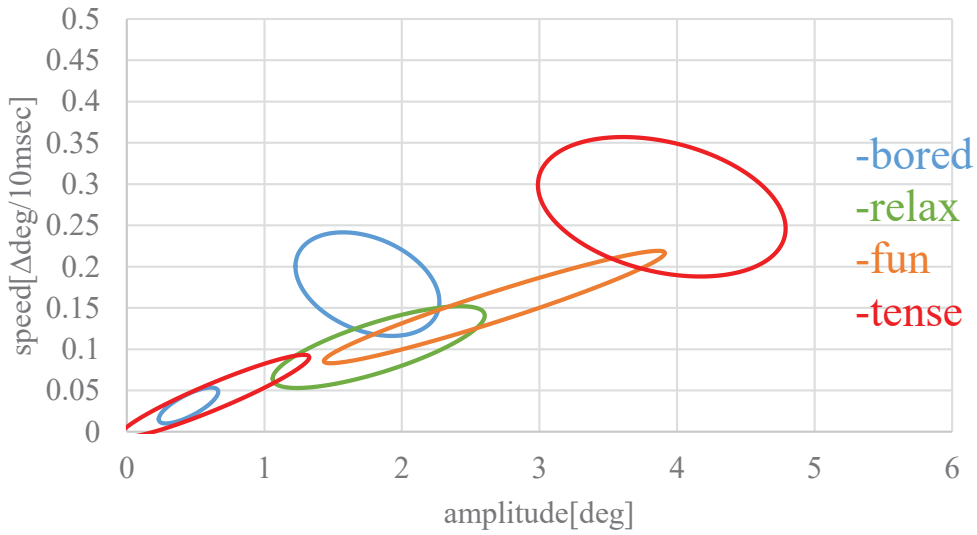


図 5.6 動き特徴空間における感情の分布

点の感情を指定すれば対応するパラメータが決まり，その感情を表現するように発話動作を変調することができる．また，感情と動作生成パラメータが連続的に対応していることで，ある感情から別の感情に中間感情を経ながら徐々に変化させた場合に，動き方も連続的に徐々に変化させることができる．このマッピングを用いることで，感情のわず

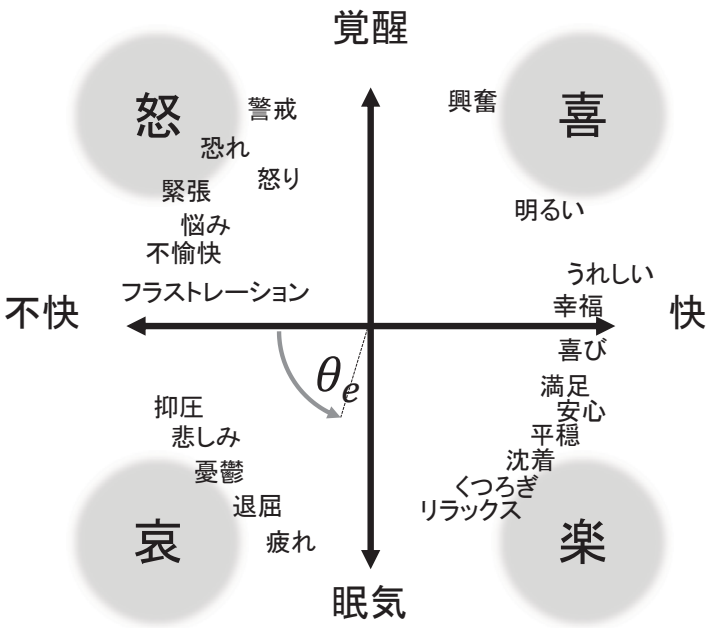


図 5.7 動き特徴と Russell の感情モデル上の感情状態との関係

かな変化を動作のわずかな変化によって表現可能なシステムを構築することができる．

表 5.1 と図 5.9 に，パラメータ調節に要する時間（単位は秒）と満足度を示す．満足度は 1 が不満，4 がどちらでもない，7 が満足である．本実験で使用した音声が 1 分程度であるため，被験者は 3 ~ 5 回程度，発話動作を繰り返し見ることでパラメータを調節したことになる．他の手法との比較結果ではないが，比較的手間をかけずに調節できていたと言える．また，満足度についてはいずれも 4 以上であり，被験者は自分が想定する動きをアンドロイドに実装できたと感じている，すなわち調節した動きに指定された感情を感じていると推測される．本実験では，いずれの感情に合わせて動作を調節する場合にも，全て同じ音声を用いた．そのため，感情と話し方が一致せず，動きを調節しにくいと報告する被験者もいた．しかし上記の結果から，被験者が主観的に感じる感情と動作のマッピングが行われていることが確認できる．また図 5.10 には，アンドロイドが「（省略）万能調味料，醤油が選ばれました.」という文章を読み上げている様子を示す．図中のグラフは頭部角度を表し，特徴的な姿勢を写真で示している．図 5.10 からわかるように，ほとんど頭部を動かさないことで退屈を表現している．また，本手法は母音が”あ”,”え”,”お”の音を発声する際に，頭部が大きく動くモデルになっている [70] が，楽

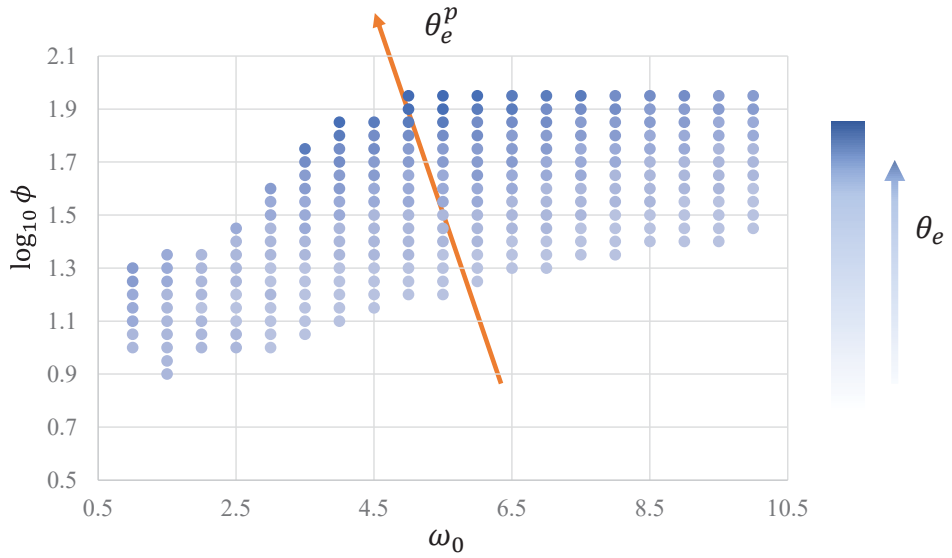


図 5.8 感情と動作生成パラメータとの関係

表 5.1 動作調節に要した時間 [秒]

	楽しい	退屈	リラックス	緊張
平均	305	294	188	278
標準偏差	205	266	132	213

しいやリラックスの感情は，文の切れ目やこれらの母音の発声に合わせてリズムよく動くことで表現されている．さらに，楽しいやリラックスよりも大袈裟に動くことで，異常な心理状態である緊張が表現されている．このように被験者が意図した動作が表現されていると考えられる．

前述したように，被験者が調節した動きはその人の個性の影響を受けている可能性がある．そのことを調べるため，NEO-FFI を用いて性格診断を行った結果から，各被験者の BigFive（神経症傾向 N，外向性 E，開放性 O，調和性 A，誠実性 C）を計算した．図 5.11 に被験者ごとの BigFive を示す．この図から被験者間で際立った性格の偏りが見られず，本実験で被験者が調節して得られた動き特徴は，被験者の BigFive 個性には存しないものであることが確認された．

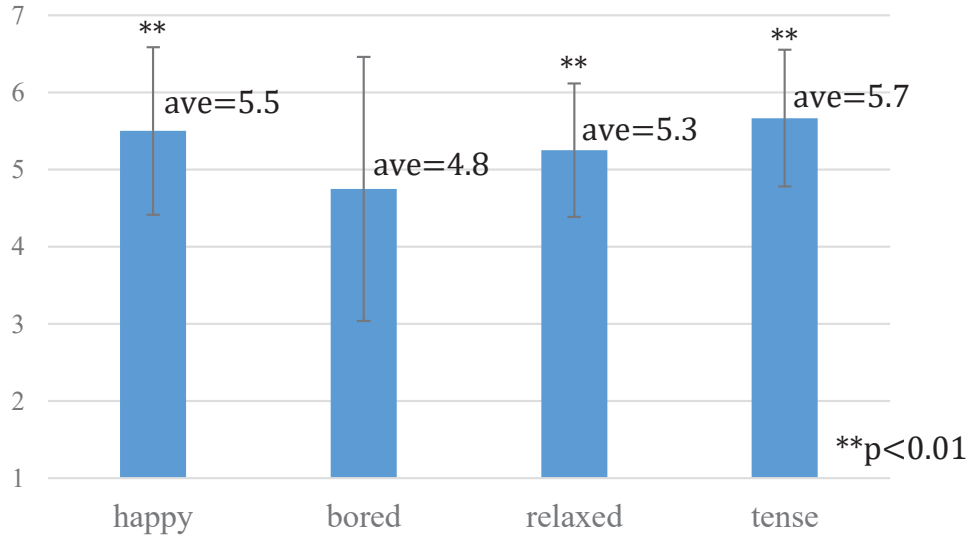


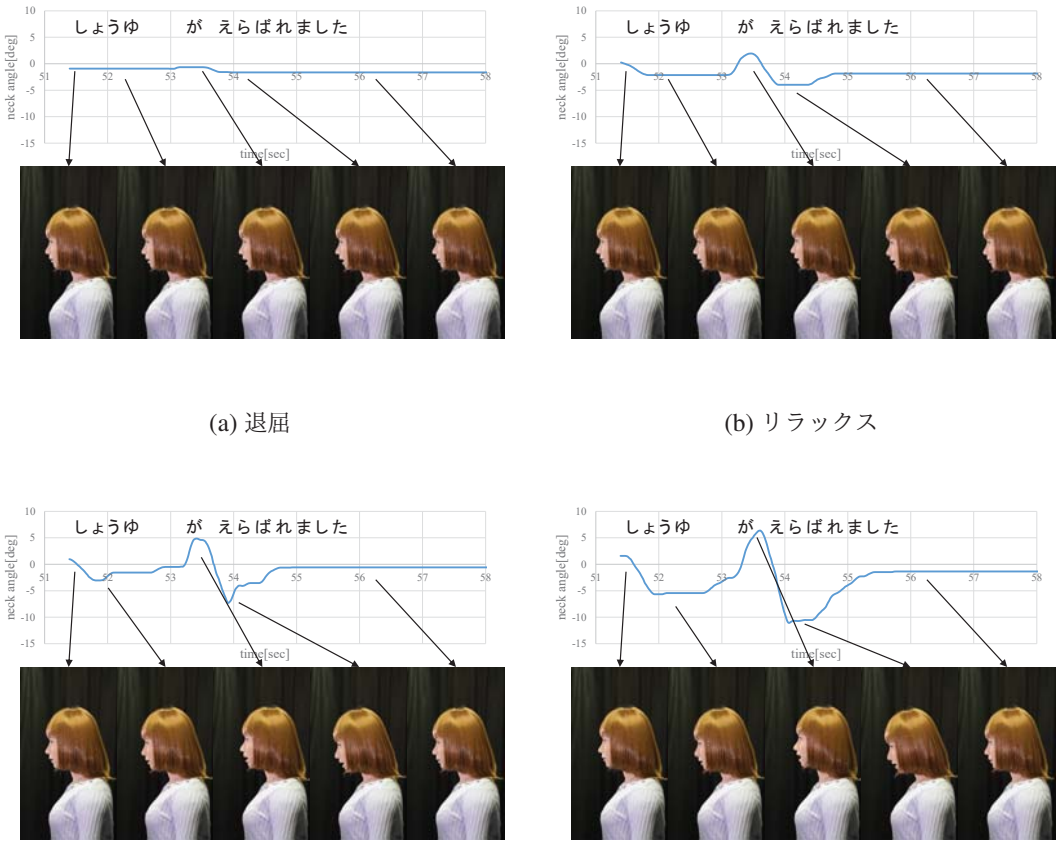
図 5.9 調節した動作に対する満足度

5.4　考察

様々な感情を表出するような動き方が，動き特徴空間上で直線状に分布した原因については，次のように考えられる．図 5.3 から分かるように，頭部は発話に合わせてリズミックな動作を行う．このときの周期は音声の韻律特徴と同期していないと不自然になるため，同じ音声で頭部の振幅を大きくすると被験者は動きが速くなるように調節し，小さくすると動きが遅くなるように調節すると考えられる．すなわち，同じ発話動作に対して動きを変調する場合，動きの大きさと速さは独立ではなく，速さは感情に適した動きの大きさと発話速度から決定される．したがって，直線状のマッピング結果は，発話動作に特有の結果である可能性がある．他の無意識的動作やジェスチャにおいても，本結果のような単純なマッピングが可能かどうかについては，今後の課題である．

本実験結果では，緊張状態を表現する動きが 2 群に分かれてマッピングされたため，感情のトーラス状の遷移が動き特徴空間上でもトーラス状の遷移として再現される結果となった．従来研究では，Russell の第 2 象限の感情動作の特徴として，今回の実験結果の緊張の 2 群のうち，動きが大きい群のみ報告されている．それらの研究では，動作と感情の関係を調べる際に，人に感情に応じた動きを演じさせ，その動きを解析するが，演者が

分かりやすい動作として，大きな動きを表現する傾向にあったためだと考えられる．本実験では，自分で動きを演じる場合とは異なり，動作生成システムのパラメータを調節することで様々な動きを客観的に探索することができる．客観的に動きを探索することで，従来の研究のように被験者のステレオタイプとは異なる感情的な動きを見つけることができ，大きな動きの緊張動作だけでなく，小さな動きの緊張動作も見つかったと考えられる．これにより，Russell の円環状の感情の連続的な遷移が，動作空間状の 1 次元に，トーラス状の性質も含めてマッピングできた．Russell の感情モデルを用いた自律対話システムはいつくか提案されており [114]，本システムはそれらのシステムと容易に組み合わせることが可能である．

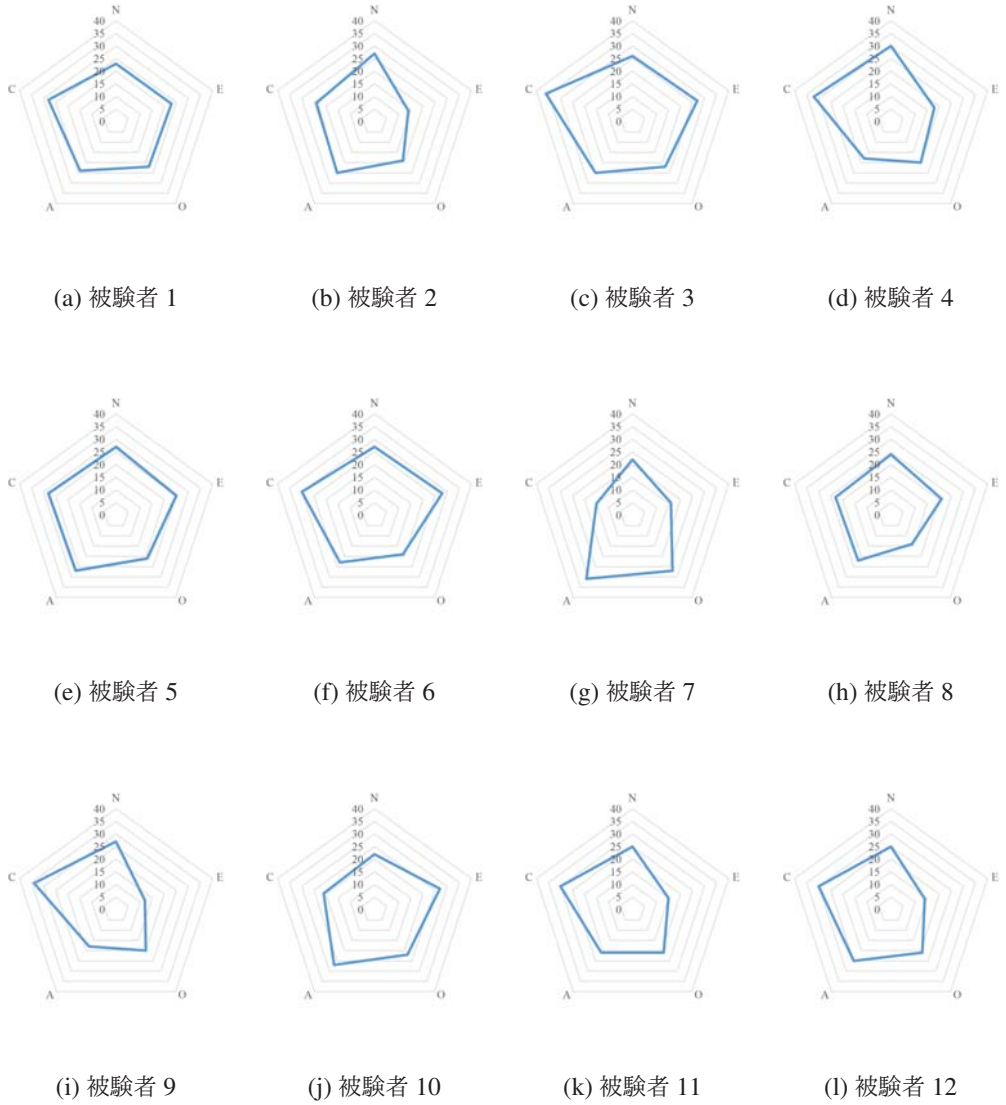


(a) 退屈　(b) リラックス

(c) 楽しい　(d) 緊張

図 5.10 感情ごとに調節された動き

本章では，調節された動きの特徴として統計量（動作中の大きさと速さの中央値）を用いたため，互いに軌跡の異なる動作が似た動き特徴を有することがある．図 5.8 の結果は，同一の感情に対応するパラメータ群を式 5.2.1 に基づいて計算したものであるが，異なるパラメータでは動作軌跡が異なるため，異なる感情を感じる可能性もある．図 5.8 の結果が被験者の主観とどの程度一致しているかを調査することが今後必要である．



(a) 被験者 1　(b) 被験者 2　(c) 被験者 3　(d) 被験者 4

(e) 被験者 5　(f) 被験者 6　(g) 被験者 7　(h) 被験者 8

(i) 被験者 9　(j) 被験者 10　(k) 被験者 11　(l) 被験者 12

図 5.11 BigFive に基づく被験者の性格診断結果

本実験結果は，単一の音声を用いた結果であり，パラメータ空間上にマッピングされる θ_e^p 軸の位置は，話し方によって変化する．特に発話速度に依存し，発話速度が速くなると θ_e^p 軸は ω_0 が大きくなる方向にシフトすると考えられる．さらに話し方は感情によっても変化する（例えば，緊張した状態では大声で発話速度が速くなり，リラックスした状態では適度な音量で発話速度は遅くなる）．これは今回得られた感情と動き特徴との関係と類似しており，動き方も話し方も本質的に感情状態が独立変数となり，それに応じて大きさ（動きの大きさ・音圧）と速さ（動きの速さ・話す速さ）が変化すると考えられる．この仮説が正しければ，発話速度を推定すれば，音声に合わせて発話に含まれる感情状態に合った動作が自動で生成できると考えられる．

本実験では成人女性型アンドロイドを使用したが，感情に適合する動作変調はアンドロイドの見かけ（年齢や性別）に依存する可能性がある．例えば男性の見かけであれば，感情表出の動きがより大きくなることが想像される．アンドロイドの外見や声色などの特徴が，感情に対応する動き変調にどのように影響するかを明らかにすることは，手法の実用性の上でも重要な課題である．

5.5　まとめ

本章では，アンドロイドの感情的な発話と一致したわずかな感情や態度の変化を，動作の変化によって表現可能な発話動作生成システムの構築を目指し，第 1 段階として第 3 で提案した音声駆動頭部動作生成システムを用いて，感情的な動き特徴，パラメータ空間と感情空間の対応を実験により明らかにした．アンドロイドの発話動作と，その動作から被験者が感じる感情との対応関係を調べた結果，動作を変調するパラメータ空間の 1 次元上の変化が，Russell の感情モデルの円周方向の変化に対応することがわかった．この対応関係を用いると，感情の細かな変化を表現するように動作を変調することができる発話動作生成システムを構築することができる．動作生成モデルのパラメータ変更によって状況に応じた動作を生成できることを示したことは，本論文の動作生成モデル構築のアプローチの有効性を示すものと考えられる．

本章の結果は，発話動作におけるものであったが，他の身体動作においても同様のシステムが構築できるかを調べることは最も重要な課題である．さらに，動作によって細かな感情変化を表出することで，人とアンドロイドのインタラクションがどのように向上するかを確かめる実験も今後必要である．

人間に酷似するアンドロイドは，細かな感情の変化を表現することで，非人型ロボットや非ロボットメディアを用いた対話よりも遙かに親和的で自然な対話を実現できるポテンシャルがあるにもかかわらず，従来研究は，表情・ジェスチャの表現方法や発話生成手法に留まっている．本章のようなシステムを用いれば，細かな感情の違いを表出することで，雰囲気の形成などこれまでにできなかった人間らしいインタラクションの実現が期待される．本章の結果は，アンドロイドの対話メディアとしての性能や意義の飛躍的な向上に貢献すると考えられる．

展望として，感情的な喋り方と感情的な動きをどのようなタイミングで同期させると効果的に人らしい感情を表現できるかを明らかにし，感情レベルで発話と動きが一致する動作生成システムの構築が課題である．

第 6 章

結論

アンドロイドでは円滑な対話を実現するために，人間らしい身体動作の実現が不可欠である．アンドロイドの身体構造やアクチュエータの動特性は人間の筋骨格系とは異なる制約があるため，アンドロイドで人間の動作を完全に再現することはできず，人間と同じ動作を行わせようとすると不自然さが残る．そのため，外見や動きなどの単一モダリティを人間らしくするだけでは不十分である．

解決策として，マルチモーダルの同期性，誇張による確証バイアスの誘発によってハードウェアの制約のあるアンドロイドで人間らしい動作を知覚させる方法を提案した．本研究では，対話する上で必ずアンドロイドが行う発話と動きの同期性に着目し，発声に伴う症状的非言語行動（第 3 章），発話の意味と一致するジェスチャー（第 4 章），感情的発話と一致した症状的非言語行動（第 5 章）の 3 つの同期に着目し，発声に伴う症状的非言語行動と発話の意味と一致するジェスチャーは自動生成システムの構築を行い，感情的発話と一致した症状的非言語行動については感情的な動作特徴を明らかにした．

これらの評価実験結果から，多様なモダリティにおいて同期性がアンドロイドの人間らしさや自然さを向上させることが示された．特に本論文では，発声しやすい身体動作，発話の意味と動きの意味の一致がアンドロイドを人間らしいと知覚させる上で重要であることが判った．このことは，上述した人らしさを知覚させるための基本的な仮説が正しいことを示すものと考えられる．

人間らしい動きは，外見にかかわらず人型エージェントに対する親密度を向上させることが報告されている [56]．その為，本研究成果はアンドロイドの自然さを向上させるだけではなく，機械らしいヒューマノイドロボットの人間らしさの向上も期待できる．

本論文では，何人かの発話と動きの同期特徴の平均をとり，一般的・平均的な人の動作生成システムの構築を行うことにより，発話に合わせ人らしく動くアンドロイドの実装が可能となった．しかし，アンドロイドが実際に社会に出て様々な役割を果たすためには，単に人らしいだけではなく，人の個性が不可欠となる．例えば，人によっては，活発な人と話すのが好きであったり，逆におとなしい人と話すのが好きであったり，病院では安心感や信頼感を感じられる必要があり，教職員には権威を感じられなければならない．そのため，状況やコンテキストを表出する個性レベルの同期性を見せることで，単なる人らしさを感じるアンドロイドから，インタラクションする個人やアンドロイドの役割や状況に適したアンドロイドを実現することが今後の大きな課題である．

研究業績

学術雑誌等（紀要・論文集等も含む）（以下は全て査読あり）

1. 境くりま, 港隆史, 石井カルロス寿憲, 石黒浩, ”わずかな感情変化を表現可能なアンドロイド動作の生成モデルの提案”, 電子情報通信学会和文論文誌 *D*, vol. J100-D, no. 3, pp. -, March, 2017.(印刷前)
2. 境くりま, 石井カルロス寿憲, 港隆史, 石黒浩, ”音声に対応する頭部動作のオンライン生成システムと遠隔操作における効果”, 電子情報通信学会和文論文誌 *A*, vol. J99-A, no. 1, pp. 14-24, January, 2016.
3. Kurima Sakai, Hidenobu Sumioka, Takashi Minato, Shuichi Nishio, and Hiroshi Ishiguro, “Motion Design of Interactive Small Humanoid Robot with Visual Illusion”, *International Journal of Innovative omputing, Information and Control*, Vol.9, No.12, pp.4725-4736, Dec. 2013.
4. 港隆史, 境くりま, 西尾修一, 石黒浩, ”運動錯視を利用した携帯型遠隔操作ヒューマノイドの運動表現”, ヒューマンインタフェース学会論文誌, vol. 15, no. 1, pp. 51-62, February, 2013.
5. Kaiko Kuwamura, Kurima Sakai, Takashi Minato, Shuichi Nishio, Hiroshi Ishiguro, ”Hugvie: communication device for encouraging good relationship through the act of hugging”, *Lovotics*, vol. 1, no. 104, February, 2014.

国際会議における発表（以下は全て口頭発表かつ査読あり）

1. Kurima Sakai, Takashi Minato, Carlos T. Ishi, Hiroshi Ishiguro, ”Speech Driven Trunk Motion Generating System Based on Physical Constraint”, *In the IEEE International Symposium on Robot and Human Interactive Communication*, Teachers

College, Columbia University, USA, pp. 232-239, August, 2016.

2. Kurima Sakai, Carlos T. Ishi, Takashi Minato, Hiroshi Ishiguro, ”Online speech-driven head motion generating system and evaluation on a tele-operated robot”, *In IEEE International Symposium on Robot and Human Interactive Communication*, Kobe, Japan, pp. 529-534, August, 2015.
3. Junya Nakanishi, Hidenobu Sumioka, Kurima Sakai, Daisuke Nakamichi, Masahiro Shiomi, Hiroshi Ishiguro, ”Huggable Communication Medium Encourages Listening to Others”, *In 2nd International Conference on Human-Agent Interraction*, Tsukuba, Japan, pp. pp 249-252, October, 2014.
4. Kaiko Kuwamura, Kurima Sakai, Takashi Minato, Shuichi Nishio, and Hiroshi Ishiguro, “Hugvie: A medium that fosters love”, *In IEEE International Symposium on Robot and Human Interactive Communication*, Aug.2013.
5. Hidenobu Sumioka, Takashi Minato, Kurima Sakai, Shuichi Nishio, Hiroshi Ishiguro, ”Motion Design of an Interactive Small Humanoid Robot with Visual Illusion”, *In The 10th Asia Pacific Conference on Computer Human Interaction*, Matsue, Japan, pp. 93-100, August, 2012.

国内学会・シンポジウム等における発表（以下は全て口頭発表かつ査読あり）

1. 境くりま, 港隆史, 西尾修一, 石黒浩, “LED 点滅による運動錯視を用いた携帯型アンドロイドの運動錯覚の生成”, HAI シンポジウムプロシーディングス, 1B-1, 京都工芸繊維大学, Dec. 2011.
2. 桑村海光, 境くりま, 港隆史, 西尾修一, 石黒浩, “遠隔コミュニケーションにおける抱擁の効果”, HAI シンポジウムプロシーディングス, 1B-3, 京都工芸繊維大学, Dec. 2012.

国内学会・シンポジウム等における発表（以下は全て口頭発表かつ査読なし）

1. 境くりま, 港隆史, 石井カルロス寿憲, 石黒浩, ”身体的拘束に基づく音声駆動体幹動作生成システム”, 第 43 回 人工知能学会 AI チャレンジ研究会, 慶応大学日吉キャンパス 来往舎, 神奈川, November, 2015.
2. 境くりま, 石井カルロス寿憲, 港隆史, 石黒浩, “発話者の音声に対応する動作生成と遠隔操作ロボットへの動作の付加効果”, 第 39 回人工知能学会 AI チャレンジ研

究会 (SIG-Challenge), 京都大学, 京都, pp. 7-13, March, 2014.

3. 中西惇也, 住岡英信, 境くりま, 中道大介, 桑村海光, 石黒浩, ”聞く力を引き出す Human-robot Intimate Interaction”, 第 32 回日本ロボット学会学術講演会（RSJ2014）, 九州産業大学（福岡）, September, 2014.

特許等（以下は出願中）

1. 境くりま，港隆史，イシイ　カルロス　トシノリ，石黒浩，アンドロイドロボットの制御システム、装置、プログラムおよび方法，特願:2015-220049.
2. 境くりま, 港隆史, 西尾修一, 石黒浩, コミュニケーション装置, 特開:2013-212300, 公開発効日 2013 年 10 月 17 日.

その他（受賞歴等）

1. 港隆史, Dylan F. Glas, Jani Even, Florent Ferreri, 境くりま, ” 人と自然に対話する自律アンドロイドの研究のためのプラットフォーム ERICA の開発”, 平成 27 年度 ATR 表彰 優秀研究賞
2. 日本科学未来館常設展示「アンドロイド―人間って、なんだ？」

国際ワークショップにおけるポスター発表

1. Kurima Sakai, Takashi Minato, Shuichi Nishio, and Hiroshi Ishiguro, “Motion Design of an Interactive Small Humanoid Robot with Visual Illusion”, *International Symposium on Cognitive Neuroscience Robotics*, Paul G. Allen Center of CSE, University of Washington, 2013.

投稿中 学術雑誌等（紀要・論文集等も含む）（以下は全て査読あり）

1. Kurima Sakai, Takashi Minato, Carlos T. Ishi, Hiroshi Ishiguro, ”Novel Speech Motion Generation by Modelling Dynamics of Human Speech Production”, *Frontiers in Robotics and AI*.

参考文献

1) J. Goetz, S. Kiesler, and A. Powers. Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *Robot and Human Interactive Communication*, pp. 55 – 60, 2003.
2) Akanksha Prakash and Wendy a. Rogers. Why Some Humanoid Faces Are Perceived More Positively Than Others: Effects of Human-Likeness and Task. *International Journal of Social Robotics*, Vol. 7, No. 2, pp. 309–331, 2014.
3) Yutaka Kondo, Kentaro Takemura, Jun Takamatsu, and Tsukasa Ogasawara. A gesture-centric android system for multi-party human-robot interaction. *Journal of Human-Robot Interaction*, Vol. 2, No. 1, pp. 133–151, 2013.
4) 中村雅巳, 松崎辰夫. 4 ヶ国語を操る接客ロボット・アクトロイド. 日本ロボット学会誌, Vol. 24, No. 2, pp. 159–161, 2006.
5) Miki Watanabe, Kohei Ogawa, and Hiroshi Ishiguro. Can Androids Be Salespeople in the Real World? In *ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 781–788, 2015.
6) Masahiro Yoshikawa, Yoshio Matsumoto, Masahiko Sumitani, and Hiroshi Ishiguro. Development of an android robot for psychological support in medical and welfare fields. In *Robotics and Biomimetics*, pp. 2378–2383, 2011.
7) Takuya Hashimoto and Hiroshi Kobayashi. Study on natural head motion in waiting state with receptionist robot SAYA that has human-like appearance. In *Robotic Intelligence in Informationally Structured Space*, pp. 93–98, 2009.
8) Takanori Komatsu and Seiji Yamada. Adaptation gap hypothesis: How differences between users' expected and perceived agent functions affect their subjective impression. *Journal of Systemics, Cybernetics and Informatics*, Vol. 9, No. 1, pp. 67–74, 2011.

9) 船越孝太郎, 小林一樹, 中野幹生, 山田誠二, 北村泰彦, 辻野広司. Artificial Subtle Expression としての明滅光源による音声対話の円滑化 (インタラクションデザイン,<特集>人とエージェントのインタラクション論文). 電子情報通信学会論文誌, Vol. 92, No. 11, pp. 818–827, 2009.
10) 小林一樹, 船越孝太郎, 小松孝徳, 山田誠二, 中野幹生. ASE に基づく相槌によるロボットとの対話体験の向上. 人工知能学会論文誌, Vol. 30, No. 4, pp. 604–612, 2015.
11) 渡辺富夫. 音声対話システムにおけるヒューマン・インタフェース : 引き込みを中心として. 情報処理学会研究報告. HI, ヒューマンインタフェース研究会報告, Vol. 96, No. 21, pp. 27–32, 1996.
12) 喜多壮太郎. ジェスチャー: 考えるからだ. 身体とシステム. 金子書房, 2002.
13) Mehrabian Albert. *Silent messages.* Oxford, England: Wadsworth, 1971.
14) Kevin G. Munhall, Jeffery A. Jones, Daniel E. Callan, Takaaki Kuratate, and Eric Vatikiotis-Bateson. Visual prosody and speech intelligibility: head movement improves auditory speech perception. *Psychological Science*, Vol. 15, No. 2, pp. 133–137, 2004.
15) Mathilde M. Bekker, Judith S. Olson, and Gary M. Olson. Analysis of gestures in face-to-face design teams provides guidance for how to use groupware in design. In *Proceedings of the 1st conference on Designing interactive systems: processes, practices, methods, & techniques*, pp. 157–166. ACM Press, New York, New York, USA, 1995.
16) Martha W Alibali, Dana C Heath, and Heather J Myers. Effects of Visibility between Speaker and Listener on Gesture Production: Some Gestures Are Meant to Be Seen,. *Journal of Memory and Language*, Vol. 44, No. 2, pp. 169–188, 2001.
17) 真介毛利, 左紀子吉川. 発話に伴う自発的身振りのコミュニケーション機能についての実験的検討 : 話し手も聞き手も注意を向けない身振りは情報伝達に寄与するか?(「手」及びヒューマン情報処理一般). 電子情報通信学会技術研究報告, Vol. 105, No. 358, pp. 131–136, oct 2005.
18) Sarah Jessen and Sonja a. Kotz. Affect differentially modulates brain activation in uni- and multisensory body-voice perception. *Neuropsychologia*, Vol. 66, pp. 134–143, 2015.
19) Katherine Isbister and Clifford Nass. Consistency of personality in interactive charac-

ters: verbal cues, non-verbal cues, and user characteristics. *International Journal of Human-Computer Studies*, Vol. 53, No. 2, pp. 251–267, 2000.

20) 山本眞理子, 原奈津子. 他者を知る: 対人認知の心理学. セレクション社会心理学. サイエンス社, 2006.

21) Gene Ball and Jack Breese. Relating personality and behavior: posture and gestures. In Ana Paiva, editor, *Affective Interactions*, Vol. 1814 of *Lecture Notes in Computer Science*, pp. 196–203. Springer-Verlag, 2000.

22) Joseph D. Matarazzo, Arthur N. Wiens, George Saslow, Bernadene V. Allen, and Morris Weitman. Interviewer Mm-Hmm and interviewee speech durations. *Psychotherapy: Theory, Research & Practice*, Vol. 1, No. 3, p. 109, 1964.

23) Machiko Sannomiya, Atsuo Kawaguchi, Ikue Yamakawa, and Yusuke Morita. Effect of backchannel utterances on facilitating idea-generation in Japanese think-aloud tasks. *Psychological reports*, Vol. 93, No. 1, pp. 41–46, 2003.

24) Lucas Kovar, Michael Gleicher, and Frédéric Pighin. Motion graphs. *ACM Transactions on Graphics*, Vol. 21, No. 3, pp. 473–482, 2002.

25) 松下善則, 川本真一, 中井満, 下平博, 嵯峨山茂樹. 擬人化音声対話エージェントにおける発話時の頭部挙動モデル. 電子情報通信学会技術研究報告, Vol. 101, No. 699, pp. 9–16, 2002.

26) Carlos Busso, Zhigang Deng, Ulrich Neumann, and Shrikanth Narayanan. Natural head motion synthesis driven by acoustic prosodic features. *Computer Animation And Virtual Worlds*, Vol. 16, No. 3-4, pp. 283–290, 2005.

27) Mary Ellen Foster and Jon Oberlander. Corpus-based generation of head and eyebrow motion for an embodied conversational agent. *Language Resources and Evaluation*, Vol. 41, No. 3-4, pp. 305–323, 2007.

28) Mehmet Emre Sargin, Yucel Yemez, Engin Erzin, and Ahmet Murat Tekalp. Analysis of head gesture and prosody patterns for prosody-driven head-gesture animation. In *Pattern Analysis and Machine Intelligence*, Vol. 30, pp. 1330–1345, 2008.

29) Binh Huy Le, Xiaohan Ma, and Zhigang Deng. Live Speech Driven Head-and-Eye Motion Generators. *Visualization and Computer Graphics*, Vol. 18, No. 11, pp. 1902–1914, 2012.

30) Soroosh Mariooryad and Carlos Busso. Generating human-like behaviors using joint,

speech-driven models for conversational agents. *Speech and Language Processing*, Vol. 20, No. 8, pp. 2329–2340, 2012.

31) Yu Ding, Catherine Pelachaud, and Thierry Artières. Modeling multimodal behaviors from speech prosody. In *Intelligent Virtual Agents*, Vol. 8108, pp. 217–228. 2013.

32) Sergey Levine, Philipp Krähenbühl, Sebastian Thrun, and Vladlen Koltun. Gesture controllers. *ACM Transactions on Graphics*, Vol. 29, No. 4, pp. 1–11, 2010.

33) Elif Bozkurt, Shahriar Asta, Serkan Özkul, Yücel Yemez, and Engin Erzin. Multimodal analysis of speech prosody and upper body gestures using hidden semi-Markov models. In *Acoustics, Speech and Signal Processing*, pp. 3652 – 3656, 2013.

34) Adso Fernández-Baena, Raúl Montaño, Marc Antonijoan, Arturo Roversi, David Miralles, and Francesc Alías. Gesture synthesis adapted to speech emphasis. *Speech Communication*, Vol. 57, pp. 331–350, 2014.

35) 誉田雅彰. 発話動作の仕組みと発話生成モデル (聴覚・音声・言語とその障害). 電子情報通信学会技術研究報告. SP, 音声, Vol. 103, No. 749, pp. 19–24, 2004.

36) 森政弘. 不気味の谷. *Energy*, Vol. 7, No. 4, pp. 33—-35, 1970.

37) Masayuki Nakane, James Everett Young, and Neil Bruce. More Human than Human ? A Visual Processing Approach to Exploring Believability of Android Faces. In *Human-agent Interaction*, pp. 377–381, 2014.

38) Waka Fujisaki, Naokazu Goda, Isamu Motoyoshi, Hidehiko Komatsu, and Shin'ya Nishida. Audiovisual integration in the human perception of materials. *Journal of Vision*, Vol. 14, No. 4, pp. 1–20, 2014.

39) Marc O Ernst and Martin S Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, Vol. 415, No. 6870, pp. 429–433, 2002.

40) 箱田裕司, 都築誉史, 川畑秀明, 萩原滋. 認知心理学, 2010.

41) Carlos Toshinori Ishi, Chaoran Liu, Hiroshi Ishiguro, Norihiro Hagita, Intelligent Robotics, and Communication Labs. Evaluation of formant-based lip motion generation in tele-operated humanoid robots. *IROS2012*, pp. 2377 – 2382, 2012.

42) 船山智, 港隆史, 石井カルロス寿憲, 石黒浩. 遠隔操作型アンドロイドの笑い動作の付加効果. 情報処理学会関西支部大会, C-07, pp. 1–7, 大阪大学中之島センター, 大阪, 2015.

43) 中俊弥, 石田亨. 3DAgent を用いた web3D コミュニケーションにおける誇張ジェス

チャ効果の考察 (ヒューマンコンピュータインタラクション). 電子情報通信学会論文誌. D, 情報・システム, Vol. 96, No. 8, pp. 1925–1934, 2013.

44) Meeri Mäkäräinen, Jari Kätsyri, and Tapio Takala. Exaggerating Facial Expressions: A Way to Intensify Emotion or a Way to the Uncanny Valley? *Cognitive Computation*, Vol. 6, No. 4, pp. 708–721, 2014.

45) 桑原明栄, 牧野光則. CG アニメーション用誇張表現作成補助システムの提案. 芸術科学会論文誌, Vol. 2, No. 1, pp. 21–30, 2003.

46) Michael J Gielniak and Andrea L Thomaz. Enhancing Interaction Through Exaggerated Motion Synthesis. *Human-Robot Interaction*, pp. 375–382, 2012.

47) Caroline Clemens and Christoph Diekhaus. Prosodic turn-yielding cues with and without optical feedback. In *Proceedings of the SIGDIAL 2009 Conference on The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, No. September, pp. 107–110, 2009.

48) 力石武信, 中村泰, 松本吉央, 石黒浩. アンドロイドの自然な待機動作のための visual saliency モデルを用いた視線制御. ロボティクス・メカトロニクス講演会講演概要集, pp. 1P1–G15(1)”–”1P1–G15(2), 2008.

49) 岡登洋平, 加藤佳司, 山本幹雄, 板橋秀一. 韻律情報を用いた相槌の挿入. 情報処理学会論文誌, Vol. 40, No. 2, pp. 469–478, 1999.

50) Tatsuya Kawahara, Takashi Yamaguchi, Miki Uesato, Koichiro Yoshino, and Katsuya Takanashi. Synchrony in Prosodic and Linguistic Features between Backchannels and Preceding Utterances in Attentive Listening. In *Asia-Pacific Signal and Information Processing Association*, No. December, pp. 392–395, 2015.

51) Takaaki Kuratate, Kevin G Munhall, Philip Rubin, Eric Vatikiotis-Bateson, and Hani Yehia. Audio-visual synthesis of talking faces from speech production correlates. In *EUROSPEECH*, 1999.

52) Dwight Bolinger. *Intonation and Its Parts: Melody in Spoken English*. Stanford University Press, 1985.

53) Kenji Amaya, Armin Bruderlin, and Tom Calvert. Emotion from Motion. In *Graphics interface*, Vol. 96, pp. 222–229, 1996.

54) M. Melissa Gross, Elizabeth A. Crane, and Barbara L. Fredrickson. Effort-Shape and kinematic assessment of bodily expression of emotion during gait. *Human Movement*

Science, Vol. 31, No. 1, pp. 202–221, 2012.

55) Hidekazu Tamaki, Suguru Higashino, Minoru Kobayashi, and Masayuki Ihara. Reducing Speech Contention in Web Conferences. In *Applications and the Internet*, pp. 75–81, 2011.

56) Lukasz Piwek, Lawrie S McKay, and Frank E Pollick. Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition*, Vol. 130, No. 3, pp. 271–277, mar 2014.

57) Jelle Saldien, Bram Vanderborght, Kristof Goris, Michael Van Damme, and Dirk Lefeber. A motion system for social and animated robots. *International Journal of Advanced Robotic Systems*, Vol. 11, No. 1, pp. 1–13, 2014.

58) Andrew G Brooks and Ronald C. Arkin. Behavioral overlays for non-verbal communication expression on a humanoid robot. *Autonomous Robots*, Vol. 22, No. 1, pp. 55–74, 2007.

59) Miles L Patterson. 非言語コミュニケーションの統合モデルに向けて. 対人社会心理学研究, 第 7 巻, pp. 67–74, 2007.

60) Tatsuya Harada, Sou Taoka, Taketoshi Mori, and Tomomasa Sato. Quantitative evaluation method for pose and motion similarity based on human perception. In *4th IEEE/RAS International Conference on Humanoid Robots*, pp. 494–512, 2004.

61) Tomio Watanabe, Masashi Okubo, Mutsuhiro Nakashige, and Ryusei Danbara. InterActor: Speech-Driven Embodied Interactive Actor. *International Journal of Human-Computer Interaction*, Vol. 17, No. 1, pp. 43–60, 2004.

62) Per-Olof Eriksson, Hamayun Zafar, and Erik Nordh. Concomitant mandibular and head-neck movements during jaw opening-closing in man. *Journal of oral rehabilitation*, Vol. 25, No. 11, pp. 859–870, 1998.

63) Carlos Toshinori Ishi, Hiroshi Ishiguro, and Norihiro Hagita. Analysis of relationship between head motion events and speech in dialogue conversations. *Speech Communication*, Vol. 57, No. 0, pp. 233–243, 2014.

64) Carlos Toshinori Ishi, ChaoRan Liu ChaoRan Liu, H Ishiguro, and N Hagita. Head motion during dialogue speech and nod timing control in humanoid robots. In *Human-Robot Interaction*, pp. 293–300, 2010.

65) Chaoran Liu, Carlos Toshinori Ishi, H Ishiguro, and N Hagita. Generation of nodding,

head tilting and eye gazing for human-robot dialogue interaction. In *Human-Robot Interaction*, pp. 285–292, 2012.

66) H. Ogawa and T. Watanabe. InterRobot: a speech driven embodied interaction robot. *IEEE International Symposium on Robots and Human Interactive Communications*, pp. 322–327, 2000.

67) Zhiyong Wu, Helen M. Meng, Hongwu Yang, and Lianhong Cai. Modeling the expressivity of input text semantics for chinese text-to-speech synthesis in a spoken dialog system. *Audio, Speech and Language Processing*, Vol. 17, No. 8, pp. 1567–1577, 2009.

68) Hiroyasu Miwa, Kazuko Itoh, Munemichi Matsumoto, Massimiliano Zecca, Hideaki Takariobu, Stefano Roccella, Maria Chiara Carrozza, Paolo Dario, and Atsuo Takanishi. Effective emotional expressions with emotion expression humanoid robot WE-4RII. In *Intelligent Robots and Systems*, Vol. 3, pp. 2203–2208, 2004.

69) Atsushi Nakano and Junichi Hoshino. Composite conversation gesture synthesis using layered planning. *Systems and Computers in Japan*, Vol. 38, No. 10, pp. 58–68, 2007.

70) Kurima Sakai, Takashi Minato, Carlos Toshinori Ishi, and Hiroshi Ishiguro. Speech Driven Trunk Motion Generating System based on Physical Constraint. In *Robot and Human Interactive Communicationtion*, pp. 232–239, 2016.

71) James A. Russell. A circumplex model of affect. *Personality and Social Psychology*, Vol. 39, No. 6, pp. 1161–1178, 1980.

72) Jia Jia, Zhiyong Wu, Shen Zhang, Helen M. Meng, and Lianhong Cai. Head and facial gestures synthesis using PAD model for an expressive talking avatar. *Multimedia Tools and Applications*, pp. 1–23, aug 2013.

73) Gunnar Johansson. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, Vol. 14, No. 2, pp. 201–211, 1973.

74) Johannes Michalak, Nikolaus F Troje, Julia Fischer, Patrick Vollmar, Thomas Heidenreich, and Dietmar Schulte. Embodiment of sadness and depression–gait patterns associated with dysphoric mood. *Psychosomatic medicine*, Vol. 71, No. 5, pp. 580–587, 2009.

75) Rudolf von Laban. *The Mastery of Movement*. Princeton Book Co.Pub., 1988.

76) 増田恵, 加藤昇平, 伊藤英則. ラバン理論に基づいたヒューマンフォームロボットの

身体動作の動作特徴抽出と表出感情推定. 日本感性工学会論文誌, Vol. 10, No. 2, pp. 295–303, 2009.

77) 中奈央子. 心理的負荷における筋弾性 と自律神経機能への影響. 口腔病学会雑誌, Vol. 72, No. 3, pp. 209–216, 2005.

78) 宇尾野公義. 自律神経失調の臨床. 新興医学出版, 1980.

79) 山下格. 精神生理的基盤. 諏訪望, 西園昌久（編）, 心身疾患 I 現代精神医学大系 7A, pp. 37–68. 中山書店, 1979.

80) C.D. Kidd and C. Breazeal. Effect of a robot on user perceptions. *Intelligent Robots and Systems*, Vol. 4, pp. 3559–3564, 2004.

81) Susumu Tachi, Naoki Kawakami, Hideaki Nii, Kouichi Watanabe, and Kouta Minamizawa. Telesarphone: Mutual Telexistence Master-Slave Communication System based on Retroreflective Projection Technology. *SICE Journal of Control, Measurement, and System Integration*, Vol. 1, No. 5, pp. 335–344, 2008.

82) Charith Lasantha Fernando, Masahiro Furukawa, Tadatoshi Kurogi, Sho Kamuro, Katsunari Sato, Kouta Minamizawa, and Susumu Tachi. Design of TELESAR V for Transferring Bodily Consciousness in Telexistence. *International Conference on Intelligent Robots and Systems*, pp. 5112–5118, 2012.

83) Tadakazu Kashiwabara, Hirotaka Osawa, Kazuhiko Shinozawa, and Michita Imai. TEROOS: A Wearable Avatar to Enhance Joint Activities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2001–2004, 2012.

84) Daisuke Sakamoto, Takayuki Kanda, Tetsuo Ono, Hiroshi Ishiguro, and Norihiro Hagita. Android as a telecommunication medium with a human-like presence. In *Human-Robot Interaction*, pp. 193–200, 2007.

85) Shuichi Nishio, Hiroshi Ishiguro, and Norihiro Hagita. Geminoid: Teleoperated android of an existing person. In *Humanoid Robots: New Developments*, No. June, pp. 343–352. I-Tech Education and Publishing, 2007.

86) Daisuke Nakamichi, Shuichi Nishio, and Hiroshi Ishiguro. Training of telecommunication through teleoperated android “Telenoid” and its effect. In *Robot and Human Interactive Communication*, pp. 1083—-1088, 2014.

87) Sean Andrist, Xiang Zhi Tan, Michael Gleicher, and Bilge Mutlu. Conversational Gaze Aversion for Humanlike Robots. In *Proceedings of the 2014 ACM/IEEE International*

Conference on Human-robot Interaction, pp. 25–32, 2014.

88) Randy J Larsen and Todd K Shackelford. Gaze avoidance: Personality and social judgments of people who avoid direct face-to-face contact. *Personality and Individual Differences*, Vol. 21, No. 6, pp. 907–917, 1996.

89) Hani C. Yehia, Takaaki Kuratate, and Eric Vatikiotis-Bateson. Linking facial animation, head motion and speech acoustics. *Journal of Phonetics*, Vol. 30, No. 3, pp. 555–568, 2002.

90) Michihiro Shimada and Hiroshi Ishiguro. Motion Behavior and its Influence on Human-likeness in an Android Robot. In *Annual meeting of the Cognitive Science Society*, pp. 2468–2473, 2008.

91) Masayuki Nakazawa, Takuya Nishimoto, and Shigeki Sagayama. Behavior generation for spoken dialogue agent by dynamical model. In *Human-Agent Interaction Symposium(in Japanese)*, pp. 2C–1, 2009.

92) Cho-chung Liang and Chi-feng Chiang. A study on biodynamic models of seated human subjects exposed to vertical vibration. *International Journal of Industrial Ergonomics*, Vol. 36, pp. 869–890, 2006.

93) Astrid Linder. A new mathematical neck model for a low-velocity rear-end impact dummy: Evaluation of components influencing head kinematics. *Accident Analysis and Prevention*, Vol. 32, pp. 261–269, 2000.

94) Hamayun Zafar, Erik Nordh, and Per-Olof Eriksson. Spatiotemporal consistency of human mandibular and head-neck movement trajectories during jaw opening-closing tasks. *Experimental Brain Research*, Vol. 146, No. 1, pp. 70–76, 2002.

95) Hamayun Zafar, Erik Nordh, and Per-Olof Eriksson. Temporal coordination between mandibular and head-neck movements during jaw opening-closing tasks in man. *Archives of Oral Biology*, Vol. 45, No. 8, pp. 675–682, 2000.

96) Dylan F. Glas, Takashi Minato, Carlos Toshinori Ishi, Tatsuya Kawahara, and Hiroshi Ishiguro. ERICA: The ERATO Intelligent Conversational Android. In *Robot and Human Interactive Communicationtion*, pp. 22–29, 2016.

97) Tanya L. Chartrand and John A. Bargh. The chameleon effect. *Journal of Personality and Social Psychology*, Vol. 76, No. 6, pp. 893–910, 1999.

98) Sandra Y. Okita and Daniel L. Schwartz. Young Children'S Understanding of Animacy

and Entertainment Robots. *International Journal of Humanoid Robotics*, Vol. 03, pp. 393–412, 2006.

99) Carlos Toshinori Ishi, Hiroshi Ishiguro, and Norihiro Hagita. Analysis of prosodic and linguistic cues of phrase finals for turn-taking and dialog acts. Proceedings of The Ninth International Conference of Speech and Language Processing 2006 (Interspeech’2006-ICSLP), pp. 2006–2009, 2006.

100) Carlos Toshinori Ishi, Hiroshi Ishiguro, and Norihiro Hagita. Automatic extraction of paralinguistic information using prosodic features related to F0, duration and voice quality. *Speech Communication*, Vol. 50, No. 6, pp. 531–543, 2008.

101) Akinobu Lee, Tatsuya Kawahara, and Kiyohiro Shikano. Julius — an Open Source Real-Time Large Vocabulary Recognition Engine. In *EUROSPEECH*, pp. 1691–1694. ISCA, 2001.

102) Carlos Toshinori Ishi. Perceptually-Related F0 Parameters for Automatic Classification of Phrase Final Tones. *IEICE transactions on information and systems*, Vol. 88, No. 3, pp. 481–488, March 2005.

103) Kohei Ogawa, Shuichi Nishio, Kensuke Koda, Koichi Taura, Takashi Minato, Carlos T Ishi, and Hiroshi Ishiguro. Telenoid: Tele-presence android for communication. In *ACM SIGGRAPH 2011 Emerging Technologies*, p. 15, 2011.

104) Kaiko Kuwamura, Takashi Minato, Shuichi Nishio, and Hiroshi Ishiguro. Personality distortion in communication through teleoperated robots. In *IEEE International Symposium on Robots and Human Interactive Communications*, pp. 49–54, 2012.

105) 中道大介, 西尾修一. 遠隔操作型コミュニケーションロボットにおける頷き動作の半自律化による操作主体感への影響. 人工知能学会論文誌, Vol. 31, No. 2, pp. H–F81_1–10, 2016.

106) Kaiko Kuwamura, Ryuji Yamazaki, Shuichi Nishio, and Hiroshi Ishiguro. Elderly care using teleoperated android telenoid. In *The 9th World Conference of Gerontechnology*, Vol. 13, p. 226, 2014.

107) 鎧沢勇, 滝川啓, 大久保栄, 渡辺義郎. 衛星通信を利用した画像会議におけるエコー及び伝搬遅延の影響. 電子情報通信学会論文誌 B, Vol. 64, No. 11, pp. 1281–1288, 1981.

108) 玉木秀和, 東野豪, 小林稔, 井原雅行. 遠隔会議における発話の衝突と 精神的ストレ

スの関係. 情報処理学会研究報告. GN, Vol. 2011, No. 10, pp. 1–6, 2011.

109) 玉木秀和, 東野豪, 小林稔, 井原雅行. 音声遅延が遠隔会議中の発話衝突と精神的ストレスに与える影響. 電子情報通信学会論文誌 D, Vol. 96, No. 1, pp. 35–45, 2013.

110) Paul Ekman and Wallace V Friesen. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Nonverbal communication, interaction, and gesture*, pp. 57–106, 1981.

111) Judith a. Hall, Sarah D. Gunnery, and Susan a. Andrzejewski. Nonverbal emotion displays, communication modality, and the judgment of personality. *Journal of Research in Personality*, Vol. 45, No. 1, pp. 77–83, 2011.

112) Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.

113) Hirotogu Akaike. Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, pp. 199–213. Springer, 1998.

114) Jinseok Woo, Janos Botzheim, and Naoyuki Kubota. Verbal conversation system for a socially embedded robot partner using emotional model. In *Robot and Human Interactive Communication*, pp. 37–42, 2015.

謝　辞

大阪大学と共同研究先の ATR で過ごした 5 年間の研究生活は，私が研究者として人生を決意する重要な 5 年間でした．このような研究環境を提供してくださった大阪大学，石黒浩教授，ATR の研究者の方々には心から感謝します．研究生活の大半を過ごした ATR では，ソフトボール大会やテニス大会など研究以外の面から研究生活を支えて頂き，そのような社風を築き上げた平田康夫社長に心から感謝します．

石黒浩教授は，世界を代表するアンドロイド研究の第一人者であり，私がアンドロイドと出会う機会を作って下さった石黒先生には感謝しています．

港隆史氏には，ATR にて 5 年に渡り私の研究を監督していただきました．研究に必要な技術のみならず，研究者としての考え方を一から教えていただき，現在の研究者として私がいるのも港氏のおかげです．

石井カルロス寿憲氏には，ロボットと音声に関わる研究者の専門家として，本研究についても重要なアドバイスをしていただき感謝しています．

住岡英信氏は，フランクな性格で気軽に研究の相談にのっていただき感謝しています．

西尾修一氏は，豊富な知識を持ち，常に客観的な視点から研究に取り組む姿勢は，私の模範とする研究者の 1 人であり，共に研究できたことを誇りに思います．

共に議論し研究に取り組んだ，船山智君と中道大介君に感謝します．

実験と実験結果解析に尽力していただいた森田美香氏，谷口愛氏，本間美奈可氏，今村由紀子氏，中村正恵氏，吉岡由紀氏に感謝します．

本研究の一部は，JST(戦略的創造推進事業), ERATO, 石黒共生ヒューマンロボットインタラクションプロジェクトの一環として行われたものです．

(以　　上)