博士学位論文

発話感情表現に基づく
反応動作と音声相槌に関する研究

令和２年３月

西田麻希子

岡山県立大学大学院
情報系工学研究科

発話感情表現に基づく反応動作と音声相槌に関する研究

目次

第 1 章

序論

1.1 本研究の背景と目的

現在，音声対話システムの研究開発が盛んに行われ，音声入力システムは日常的に使用できる身近な技術である．音声操作可能なシステムとして受付ロボットなどの案内エージェントや，Apple 社の「Siri」，Google 社の「Google Home」などが実サービスとして展開されている．音声認識システムに身体エージェントをもたせた「スマートメイちゃん」［1.1］のような汎用性のあるプラットフォームも運用されつつあり，今後音声対話システムは様々な分野で発展すると期待される．しかし，現在利用されている対話システムでは，人同士の対話に存在するノンバーバル情報が十分に考慮されているとは言えず，使用者が違和感を感じることがある．

人の対面コミュニケーションでは，言葉によるバーバル情報だけではなく，うなずきや身振り・手振りといった身体動作，音声情報に付随する韻律などの周辺言語（paralanguage），視線・表情などの言葉によらないノンバーバル情報が重要な役割を果たしている［1.2］．コミュニケーションにおけるノンバーバル情報の占める割合は約 65 %［1.3］とも約 93 %［1.4］とも言われており，その重要性がわかる．

コミュニケーションにおける発話音声と身体動作の関係について，Condon らによる研究では，母子間のコミュニケーションにおいて，母親の発話リズムと乳児の身体リズムが同調することが示されている［1.5］．母親の語りかけに乳児が手足を動かし

て応えるなど, 本来別々の身体的リズムで行われている行動が相互に時系列的に関係が成立して同期する現象は引き込み現象と呼ばれ, このようなリズムの同調が円滑なコミュニケーションに重要な役割を果たしている. また情動変動と密接に関連した心拍変動の引き込みや呼吸の引き込み等, 生理的側面での生体リズムの引き込みも, コミュニケーションに重要な役割を果たしている［1.6］. これらノンバーバル情報と生体情報をも含めた身体全体を介してのコミュニケーションは身体的コミュニケーションと呼ばれるもので, 一度自己の身体を介することで相手との関係を築くコミュニケーションである. 原初的コミュニケーションである母子間のインタラクションでは, この身体的コミュニケーションが主体であり, 後に発達してくる言語によるコミュニケーションよりも本質的重要性を持っていると考えられ［1.7］, 人とのインタラクションを行う対話エージェントの研究開発においても身体的なかかわりはインタラクションに重要な役割を果たすと期待される.

また, 今後さらに音声対話システムを発展させていくには, 前述のような身体的リズムの同調に加えて使用者の状態に即した応答を行うことも重要である. 相手の感情を汲み取った応答は人同士の対話における重要な要素のひとつであり, 近年, 身体動作や情動表出による社会的相互作用に着目して, 人やロボットとのインタラクションに応用する試みも研究が進められ［1.8, 1.9］, 使用者の感情推定に基づく応答を行うことで, より使用者に寄り添った対話エージェントの開発が期待されている.

渡辺らはこれまでに, 発話音声と身体動作が同期する身体的引き込みに着目し, インタロボット技術 iRT を開発してきた. iRT は音声と身体動作の関係をモデル化することで, 発話音声の有無（音声の呼気段落区分での ON-OFF パターン）から身体的引き込みを伴ううなずきなどのコミュニケーション動作を自動生成し, コミュニケーションを支援するものである. この iRT を CG キャラクタに組み込んだ音声駆動型身体的引き込みキャラクタ InterActor を開発し, コミュニケーション支援の有効性を示している［1.10］. 従来の InterActor は, 発声の有無に基づいてうなずきなどのキャラクタの身体動作が生成されることによる効果を確認してきたが, 使用者の状態によって反応を変化させる機能は実装されていなかった. そのため, 使用者が否定的な発話をした場合においてもキャラクタがうなずき動作を行ってしまい, 使用者がうなずき

を肯定動作と解釈した場合，その否定的な発話感情を肯定し，助長してしまう可能性がある．

そこで本研究では，使用者の負の感情を助長しないという観点から，身体的引き込みモデルで動作するキャラクタに対して，特徴的な身体的リズム同調を損なうことなく，使用者の発話感情に応じた反応動作または音声相槌を自動生成するシステム開発とその評価を行う．

1.2 関連研究

近年，音声認識および音声合成技術を用いた対話エージェント等の研究開発が積極的に進められており，タスク型対話システムだけでなく，対話相手としての非タスク型（雑談型）対話システムの構築を目的とした研究開発も進められている．人間とコミュニケーションを行うロボットやエージェントの開発において，身体動作や相槌などの応答反応は重要な要素であり，適切な応答の生成を試みる様々な研究が進められている．

石井らはより親和的なキャラクタを実現するために，キャラクタらしい発話ペアデータを効率的に収集する「なりきり質問応答」という方法論を利用して，既存アニメキャラクタを用いたアニメーション付きテキスト対話システムを提案・開発し，人間の発話に対応する身体動作のデータを用いて，任意の発話文から身体動作を自動生成する手法を提案している［1.11］．

相槌の生成および挿入タイミングについても多くの検討が行われている．岡登らはテレフォンショッピングをタスクとした音声発話において，テンプレートを用いた韻律パターンの認識による相槌タイミングの検出方法を提案している［1.12］．北岡らは談話コーパスより発話タイミングを決める決定木を求め，それを用いて自然な発話タイミングで対話することができるシステムの構築を行っている［1.13］．またそのアルゴリズムを用いて，小林らはロボットやエージェントによる相槌に着目し，人間的な表現を用いない相槌生成手法として，箱型のロボットが光の明滅やビープ音を提示して相槌を表現し，それが対話体験に対して与える影響を調査している［1.14］．

音声からユーザの発話感情を推定し, 音声対話エージェントへ適用する手法についても検討が進められており, 発話区間の時間的関係性に基づく対話的特徴（発話時間, 相槌回数，発話権交替時間，発話時間比）を用いたコールセンタ対話における顧客のオペレータに対する怒り状態の推定を行う手法の提案［1.15］や，ニューラルネットワークを用いて構文解析等テキスト処理を行う Word2Vec［1.16］を用いて言語的な意味の近接性を求める手法の応用が挙げられる．さらに，松井らは感情推定機能と知識獲得機能を有する対話システムの提案として, 入力音声に対して音声処理による感情推定と, 音声認識により得られたテキスト文による自然言語処理による感情推定の 2 つの方法によりバイモーダル的感情推定を行うとともに，知識獲得では，単語のベクトル情報と関係パターンを用いて, 入力文から雑談の話題が広がるような単語群の獲得を行うことで, 雑談の流れを考慮して対話システムの構築を行っている［1.17］. また山口らは相槌の形態と先行発話の統語構造や韻律的特徴との関係について分析し，これらの先行発話の特徴から相槌の形態の予測と生成を行い，ユーザ発話の文脈に応じて適切な形態の相槌をうつことができる傾聴対話システムを提案している［1.18］．Lala らは自律型アンドロイド ERICA を用いて自然な傾聴対話を実現する研究開発を進めており，相槌のタイミングを 100 ms 毎に，その時点から 500 ms 以内に相槌を打つか否かを, 韻律情報を用いてロジスティック回帰モデルにより予測して出力することで，より早いタイミングでの相槌出力を実現している［1.19］.

非言語的なインタラクションとして重要な身体行動の同調性については，対話者の身体性の共有に着目し, 同期現象を創出する共有仮想空間を活用したシステム開発も行われている［1.20］．発話音声と身体動作の関連についても，ロボットとのコミュニケーションを対象に検討が進められている. 山本らは人の挨拶における身体動作と発声の生成タイミングの解析を行い，ロボットに実装することで，ロボットの動作開始から 300 ms 遅延した発声が自然に感じられ，より大きな発声遅延では挨拶が丁寧に感じられることを明らかにしている［1.21］．また高杉らは発話と反応潜時長間の相関関係を基にタイミング制御モデルをコミュニケーションロボットに実装し, タイミング制御の有無によって対話の印象評価に差が現れることを明らかにしている［1.22］．小林らは，発話タイミングを制御する音声対話システムにおいて，600ms

程度の固定長か, 発話長に合わせて緩やかに同調するモデルが自然かつ話しやすいという評価であったと報告している［1.23］．さらに挨拶時の話者間においては，発話リズムや身体リズムに関わる様々な時間特徴量間で同調が現れやすく, 話者内においては, 親密的関係の挨拶動作で発話リズムと身体リズムの同調がより強くなることが示されている［1.24］.

以上のようにエージェントを介したインタラクション･コミュニケーションを支援する研究開発は数多くなされている．しかしながら，身体的リズムの引き込みを基盤として, そのリズム同調を損なうことなく使用者の発話感情に応じた反応を自動生成するアプローチでの研究はなされていない.

1.3 本論文の構成

本論文は本章を含め 7 章で構成されている. 本章を除いた 2 章以降の概要を以下に述べる.

2 章では，身体的リズム同調を損なうことなく使用者の状態に寄り添った応答を行うことを目的に，音声駆動型身体的引き込みキャラクタ InterActor に音声認識を導入し，話者の発話内単語の感情極性から使用者の状態を推定し，結果に基づいて反応動作を変化させる音声駆動型身体的引き込みキャラクタシステムの開発を行った. まず, 開発したシステムのコンセプトについて記述し, 次に本研究で提案する発話感情推定方法とシステム構築について記述している.

3 章では，2 章で開発した単語感情極性に基づく音声駆動型身体的引き込みシステムの評価実験を行った．まず，Negative または Positive と判定されたシナリオに基づくキャラクタへの語りかけによる自動応答エージェントシステムとしての評価実験を行った. 次に二者対話によるコミュニケーションインタフェースシステムとしての評価実験を行い，コミュニケーション支援への有効性を示した.

4 章では，使用者の多様な感情に対応するために，2 章で開発したシステムに発話時間率に基づく活性度を加えて使用者の状態を推定する状態推定モデルを定義し, 結果に基づき反応動作を行う身体的引き込みキャラクタシステムを開発した. システム

のコンセプトとシステム構築, 開発したシステムを用いて行った発話活性度を考慮した状態推定の有効性を検討する評価実験について記述している.

5 章では，iRT によるうなずきの出力タイミングに基づく音声相槌出力について検討した．従来の InterActor には音声による応答機能は実装されていなかったが，うなずきに相当する音声相槌を意味的な解釈として出力させることで使用者の状態に対応したシステムとして活用できる可能性がある．そこで従来の InterActor に音声相槌を付加したキャラクタシステムを開発し，自動生成する音声相槌について，その出力タイミングとキャラクタ表示による効果および音声相槌の出力頻度に着目した評価実験を行い，効果を確認するとともにシステムの有効性を示した.

6 章では，5 章で開発したうなずき動作に音声相槌を伴う音声駆動型身体的引き込みキャラクタシステムにおいても, 否定的な発話感情を助長する可能性を回避する手法を提案した．まず，うなずき動作に対する音声相槌の提示タイミングを検討するための評価実験及び Negative な発話が行われた際の音声相槌の提示タイミングを検討するための評価実験について記述した．次に，得られた知見をシステム統合することで発話内単語の感情極性に基づいた音声相槌を伴う身体的引き込みキャラクタシステムの開発を行い，システムのコンセプトとシステム構築について記述している.

7 章では，本研究を通じて得られた成果を総括するとともに，今後取り組むべき課題について述べている.

参考文献

［1.1］　山本大介，大浦圭一郎，西村良太，打矢隆弘，内匠逸，李晃伸，徳田恵一，スマートフォン単体で動作する音声対話 3D エージェント「スマートメイちゃん」の開発, インタラクション 2013　IPSJ Symposium Series (2013), pp.675–680.

［1.2］　黒川隆夫，ノンバーバルインタフェース，オーム社 (1994).

［1.3］　Birdwhistell, R. L., Kinesics and context: Essays on body motion communication, University of Pennsylvania Press (1970).

［1.4］ Mehrabian, A., Communication without words, Psychology Today, Vol.2, No.4 (1968), pp.52–55.

［1.5］ Condon, W. S. and Sander, L. W., Neonate movement is synchronized with adult speech, Science, Vol.183 (1974), pp.99–101.

［1.6］ 渡辺富夫，大久保雅史，コミュニケーションにおける引き込み現象の生理的側面からの分析評価，情報処理学会論文誌, Vol.39, No.5 (1998), pp.1225–1231.

［1.7］ Kobayashi, N., Ishii, T. and Watanabe, T., Quantitative evaluation of infant behavior and mother-infant interaction: an overview of a Japanese interdisciplinary programme of research, Early Development and Parenting, Vol.1, No.1 (1992), pp.23–31.

［1.8］ アレックス (サンディ) ペントランド著，柴田裕之訳，安西祐一郎監訳，正直シグナル非言語コミュニケーションの科学, みすず書房 (2013) (原著 : Pentland, A., Honest signals: how they shape our world, The MIT Press (2010)).

［1.9］ Woo, J., Botzheim, J. and Kubota, N., Verbal conversation system for a socially embedded robot partner using emotional model, Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication (2015) pp.37–42.

［1.10］ Watanabe, T., Okubo, M., Nakashige, M. and Danbara, R., InterActor: speech-driven embodied interactive actor, International Journal of Human-Computer Interaction, Vol.17, No.1 (2004), pp.43–60.

［1.11］ 石井亮, 東中竜一郎, 水上雅博, 片山太一, 光田航, 川端秀寿, 山口絵美, 安達敬武, 富田準二, 既存のアニメキャラクタを用いたテキスト対話システム構築手法, HAI シンポジウム 2018 論文集, G–3 (2019), pp.1–11.

［1.12］ 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 (1999), pp.469–478.

［1.13］ Kitaoka, N., Takeuchi, M., Nishimura, R. and Nakagawa, S., Response Timing detection using prosodic and linguistic information for human-friendly spoken

dialog systems, Transactions of the Japanese Society for Artificial Intelligence, Vol.20, No.3 (2005), pp.220–228.

［1.14］ 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 (2015), pp.604–612.

［1.15］ 野本済央, 小橋川哲, 田本真詞, 政瀧浩和, 吉岡理, 高橋敏, 発話の時間的関係性を用いた対話音声からの怒り感情推定, 電子情報通信学会論文誌 D, Vol.J96−D, No.1 (2013), pp.15–24.

［1.16］ Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean, J., Distributed representations of words and phrases and their compositionality, Proceedings of the 26th International Conference on Neural Information Processing Systems, Vol.2 (2013), pp.3111–3119.

［1.17］ 松井辰哉, 萩原将文, 感情推定と知識獲得機能を有する対話システムの構築, 日本感性工学会論文誌, Vol.16, No.1 (2017), pp.35–42.

［1.18］ 山口貴史, 井上昂治, 吉野幸一郎, 高梨克也, Ward Nigel G., 河原達也, 傾聴対話システムのための言語情報と韻律情報に基づく多様な形態の相槌の生成, 人工知能学会論文誌, Vol.31, No.4 (2016), pp.1–10.

［1.19］ Lala, D., Milhorat, P., Inoue, K., Ishida, M., Takanashi, K. and Kawahara, T., Attentive listening system with backchanneling, response generation and flexible turn-taking, In SIG-DIAL (2017), pp.127–136.

［1.20］ Watanabe, T. and Miwa, Y., Duality of embodiment and support for co-creation in hand contact improvisation, Journal of Advanced Mechanical Design, System, and Manufacturing, Vol.6, No.7 (2012), pp.1307–1318.

［1.21］ 山本倫也, 渡辺富夫, ロボットとのあいさつインタラクションにおける動作に対する発声遅延の効果, ヒューマンインタフェース学会論文誌, Vol.6, No.3 (2004), pp.87–94.

［1.22］ 高杉将司, 山本知仁, 武藤ゆみ子, 阿部浩幸, 三宅美博, コミュニケーションロボットとの対話を用いた発話と身振りのタイミング機構の分析, 計測

自動制御学会論文集, Vol.45, No.4 (2009), pp.215–223.

［1.23］　小林弘幸, 大村卓矢, 山本知仁, 音声対話システムにおける挨拶発話の適切なタイミング生成, 計測自動制御学会論文集, Vol.51, No.4 (2015), pp.233–239.

［1.24］　杵鞭健太, 山本知仁, 挨拶行為における発話リズムと身体リズムの同調, ヒューマンインタフェース学会論文誌, Vol.18, No.4 (2016), pp.415–424.

第 2 章

単語感情極性に基づく

音声駆動型身体的引き込みシステムの開発

2.1 はじめに

人とのインタラクションを行う対話エージェントの研究開発において, 身体的なかかわりはインタラクションに重要な役割を果たすと期待されるが，従来の InterActor は音声の ON-OFF パターンに基づいて身体性の共有に特化して開発・展開されてきており，使用者の状態によって動作を変化させる機能は実装されていなかった．そのため，“学校に行くのが辛い”や“もう駄目だ”のような否定的な発話内容の場合においても，リズム同調的に自動生成されたうなずき動作を行う．うなずきは肯定的な意味を連想させるため使用者がうなずきを肯定動作と解釈した場合, その否定的な発話感情を肯定し，助長してしまう可能性がある．

そこで本章では，発話の音量のみに基づいた従来の InterActor の身体的引き込み動作に加えて，音声認識を導入して発話内単語の感情極性から使用者の感情を推定し，結果に基づいて反応動作を変化させることで, 使用者の否定的な発話を抑制し肯定的な内容へ誘導, または肯定的な状態を保持することを目的とした身体的引き込みキャラクタシステムを開発した．

2.2 コンセプト

開発したシステムのコンセプトを図 2.1 に示す. 使用者が対話エージェントとしての CG キャラクタに話しかけると, キャラクタはうなずきなどの身体的引き込み動作を行う. そこに音声認識を併用して, 発話文の単語が一般的に良い印象を持つか, それとも悪い印象を持つかを表した二値属性である感情極性から使用者の感情を推定し, それに対応した反応動作を生成する. 使用者の否定的な発話に対しては, うなずくのではなく負の感情を緩和させるために抑制を促す動作や否定的な動作を行うことで, リズム同調を保ちつつ負の感情の抑制を促す. また肯定的な発話に対しては, 正の感情を盛り上げる動作を行うことで正の感情の促進を促す. これらの機能を用いることで, 会話意欲のさらなる促進が可能となると期待される.

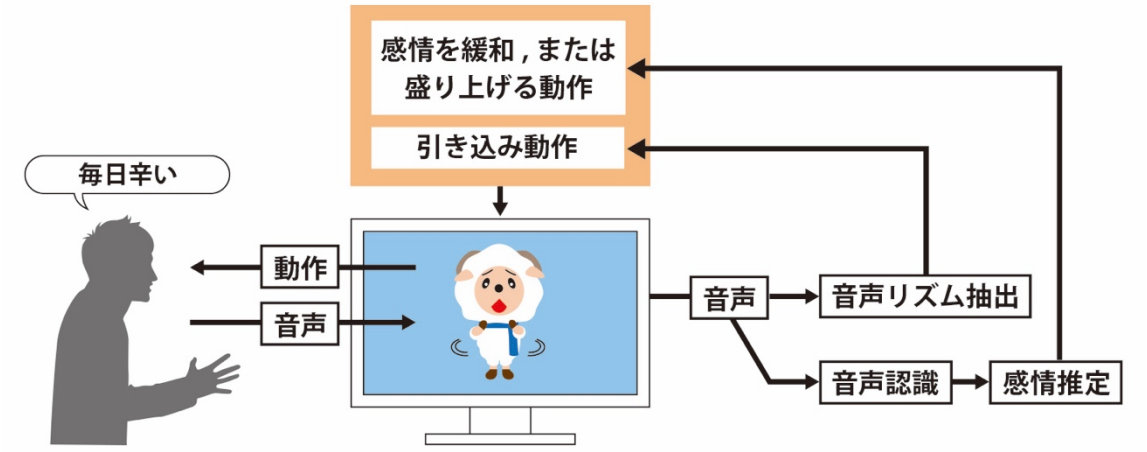

図 2.1　コンセプト

2.3 音声駆動型身体的引き込みキャラクタ InterActor

音声駆動型身体的引き込みキャラクタ InterActor は, 基本性能として音声から自動生成される身体動作による話し手と聞き手の両機能を備えている. 各関節部位の曲げ動作や回転動作を組み合わせることで, 多様なコミュニケーション動作を表現するこ

とができる（図 2.2，2.3）．図 2.4 は InterActor を用いた身体的インタラクションシステムのコンセプトである．InterActor は，聞き手として対話者の語りかけに対してうなずきや身体動作などの身体全体で引き込むように自動的に反応し，話し手として口の開閉や身体動作など話者の発話音声に同期した動作をすることで，インタラクティブなコミュニケーションを実現している．

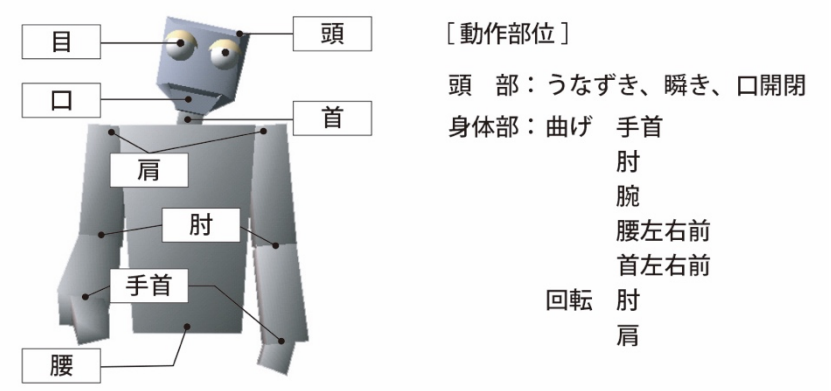

図 2.2　InterActor

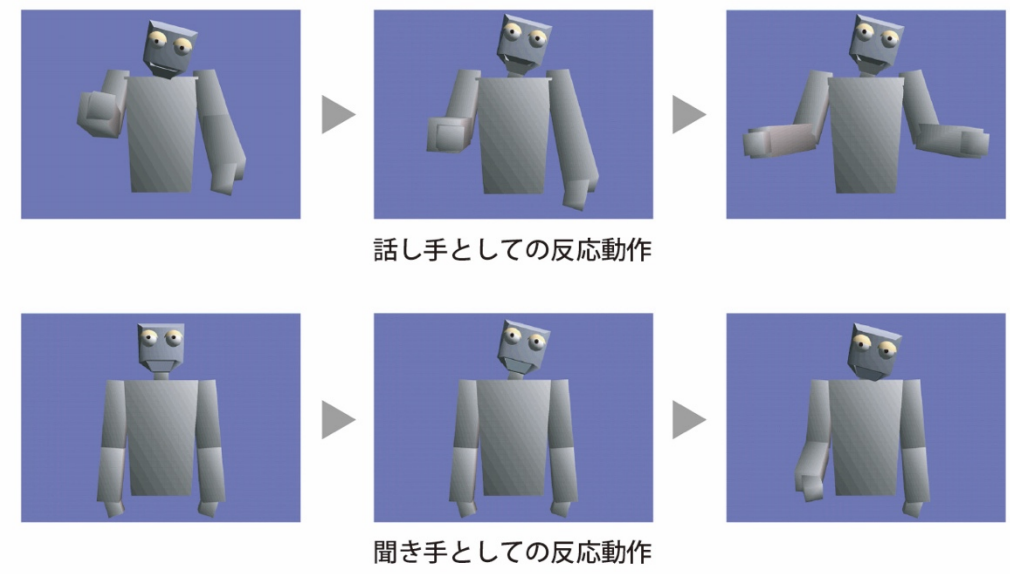

図 2.3　InterActor のコミュニケーション動作

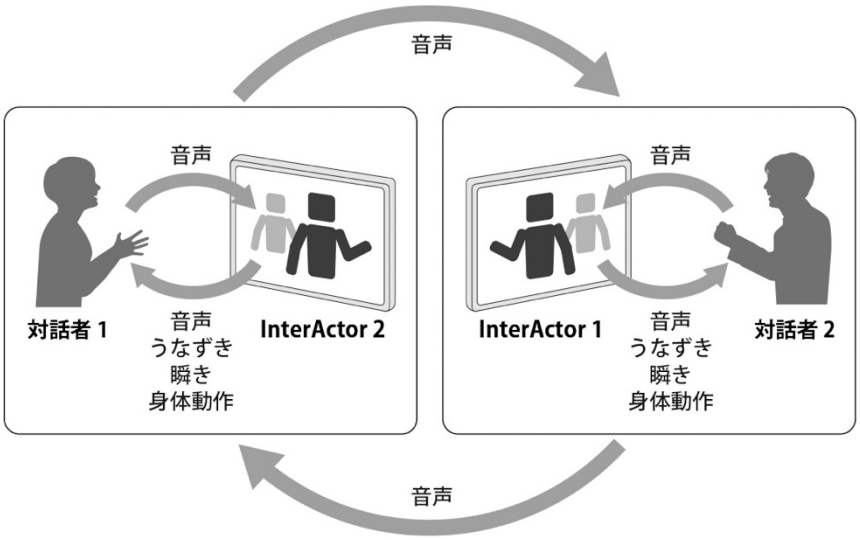

図 2.4　身体的インタラクションシステムのコンセプト

2.4 InterActor のインタラクションモデル

InterActor のインタラクションモデルを図 2.5 に示す．InterActor のインタラクションモデルは，音声の ON-OFF パターンに基づくうなずき反応モデルと閾値の異なる身体的引き込み動作の自動生成モデルで構成されている［2.1］．音声データ $V(i)$ は 16bit 22.05 kHz でサンプリングし，閾値で二値化するとともに，発声器官のメカニズム上生じる無声子音の前の短い OFF 区間による発話の断片化を除去するために 133 ms でハングオーバー処理（一定の長さの OFF 区間が続くまで ON 区間が続いているとする処理）を施している．133 ms は，秒間 30 フレームの描画の 4 フレーム分相当の時間にあたり，先行研究［2.2］において人の発話速度に対して最も相関の高い値である．このモデルでは，マクロ層とミクロ層からなる階層モデルを用いて，InterActor のうなずきの予測を行っている．マクロ層では音声の呼気段落区分での ON-OFF 区間からなるユニット区間にうなずきが存在するかを［i-1］ユニット以前のユニット時間率（ユニット区間での ON 区間の占める割合，式（2.2））の線形結合で表せる式

（2.1）の MA（Moving-Average）モデルを用いている．雑音は身体のゆらぎを示すノイズ成分である．

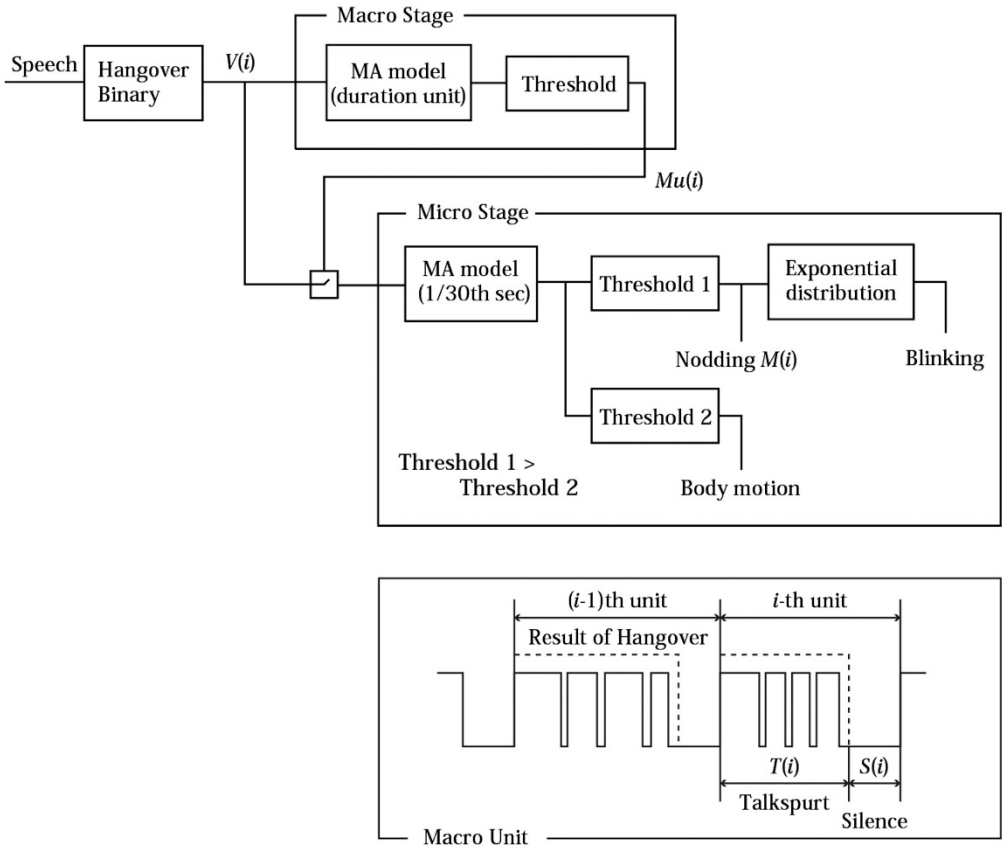

図 2.5　インタラクションモデル（聞き手）

$$Mu(i) = \sum_{j=1}^{j} a(j)R(i-j) + u(i) \tag{2.1}$$

$$R(i) = \frac{T(i)}{T(i) + S(i)} \tag{2.2}$$

$a(j)$: 予測係数

$T(i)$: i番目ユニットでの ON 区間

$S(i)$: i番目ユニットでの OFF 区間

$u(i)$: ノイズ

$$M(i) = \sum_{j=1}^{k} b(j)V(i-j) + w(i) \tag{2.3}$$

$b(j)$: 予測係数

$V(i)$: 音声データ

$w(i)$: ノイズ

予測値 *Mu*(*i*)が閾値 Threshold の 0.20 を超えて，うなずきが存在すると予測された場合は，処理はミクロ層に移る．ミクロ層では音声の ON-OFF データ（30Hz, 60 個）を入力とし，式（2.3）を用いて MA モデルでうなずきの開始時点を推定する．予測値が閾値 Threshold 1 の 0.47 を超えた場合には InterActor をうなずかせる．瞬きについては，対面コミュニケーション時における瞬き特性に基づいて，うなずきと同時に瞬きをさせ，それを起点にして次の瞬きまでの時間間隔を指数分布させている．その他の身体反応の推定にはうなずきの予測値を用い，うなずきよりも低い任意の閾値 Threshold 2 で InterActor の各部位（頭部，胴部，右肘，左肘）のうちいくつかを選択して動作させることで話し手の発話音声と関連付けている．

話し手のモデルについても同様に，対面コミュニケーション時の音声と身体動作の特性から，音声の ON-OFF パターンに基づく身体全体の動作を予測するモデルと音声の振幅に基づく腕部動作モデルを導入している．身体動作モデルとしてはすべての動きの ON-OFF の総和データから体の動くタイミングを予測し，閾値を超えたときに InterActor の各部位（頭部，胴部，右肘，左肘）のうちいくつかを選択し，動作させることで発話音声と関連付けている．

2.5 発話感情推定方法

本システムでは，使用者の発話感情推定に高村らの提案する単語感情極性対応表［2.3］を用い，発話文中の単語から正負の感情推定を行った．人間の感情区分としてはさらに細分可能であるが［2.4，2.5］，本システムでは発話のみから推定を行うた

め感情極性に着目している．単語の感情極性とは，その単語が一般的に良い印象を持つか，それとも悪い印象を持つかを表した二値属性である（以下，感情極性対応表に基づく正の値を Positive，負の値を Negative と表記）．例えば「良い」「美しい」などは Positive な極性，「悪い」「汚い」などは Negative な極性を持つ．感情極性値は語彙ネットワークを利用して自動的に計算されたものであり，-1 から+1 の実数値を割り当てている．+1 に近いほど Positive であり，-1 に近いほど Negative である．対応表には約 5 万単語が登録されているが，本システムでは極性値が 0.7 以上の単語及び-0.7 以下の単語（8351 語）を採用し，登録された単語に紐づけられた極性値に基づいて，発話文の極性値を推定した．また，対応表には単語の基本形のみが登録されているため，例えば「悲しい」という表現に対し，「悲しかった」「悲しくなる」などといった活用の場合ではマッチングできない．そこで，文章の内容を考慮するためにオープンソースの形態素解析システム MeCab を用いて品詞や内容を判別する形態素解析を行った．これにより「悲しかった」という活用の場合にも，「悲しい」という形容詞と「た」という助動詞に分割されるため，対応表に登録している「悲しい」という表現にもマッチングが可能となる. 取得した単語情報と対応表にある単語を照合し, 対応表に登録されている単語の場合には, その単語に紐付けられた値を参照する. 発話文中の単語の極性値 x_i (-1< x_i <1)，およびその単語数 n を用い，以下の式 (2.4) から発話文の感情極性値を算出した．「美しくない（美しい/ない）」といった否定語を含む場合は，「ない」の直前の単語の極性値は参照しない.

$$\text{文の感情極性値} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.4}$$

x_i：単語の極性値，n：単語数

2.6 システム構築

音声認識エンジン Julius［2.6］により，発話音声をテキストに変換し，テキスト内の単語情報から感情極性を推定して, その推定結果に応じキャラクタの動作を変化させる．推定結果が Positive の閾値を上回った場合に，正の感情を盛り上げる動作,

Negative の閾値を下回った場合に，負の感情を緩和させる動作をキャラクタの身体動作に関連付けた．Positive および Negative の閾値に関しては，予備検討として「すごく楽しかったエピソード」「自分のことが嫌になってしまったエピソード」というテーマのもと，キャラクタに対し話しかけるような口語口調で実験参加者に文章を記述させ（平仮名 400 文字程度），そこから実験参加者毎の各感情極性値の平均を算出した．さらに，各個人の平均値から全体の平均値を算出することで，Positive/Negative の閾値を決定した（表 2.1）．音声情報を用いて全ての応答動作を自動生成するため，マイクと PC のみでシステムを構成できる．Windows 7 32 bit PC（HP Core i5 CPU, 2.67 GHz, 4 GB メモリ），DirectX 及び Visual Studio で開発したプロトタイプ画面を図 2.6 に示す．

表 2.1　実験参加者毎の各感情極性値の平均と全体の平均値

Subject	1	2	3	4	5	6	7	8	9	10	Total: threshold
Negative	-0.58	-0.75	-0.92	-0.50	-0.92	-0.99	-0.73	-0.25	-0.99	-0.92	-0.76
Positive	0.56	0.86	0.85	0.65	0.97	0.58	0.58	0.99	0.65	0.52	0.72

図 2.6　システム画面

2.7　同調動作・緩和動作提示確率の決定

2.7.1　同調動作・緩和動作提示のための動作評価実験

発話文から推定した感情極性値に基づいて行う正の感情に対する強い同調および負の感情に対する緩和動作として，多様な表現を行うために，動作をそれぞれ 4 パターンずつ用意した．強い同調動作として，A：連続うなずき＋笑顔，B：深いうなずき＋笑顔，C：うなずき＋拍手＋笑顔，D：うなずき＋笑顔，緩和動作として E：手と頭を左右に振る（否定），F：腕を組む＋首をかしげる（疑問），G：口を開閉＋腕を上下（宥める），H：手を横に出す（ツッコミ）動作である（図 2.7，2.8）．

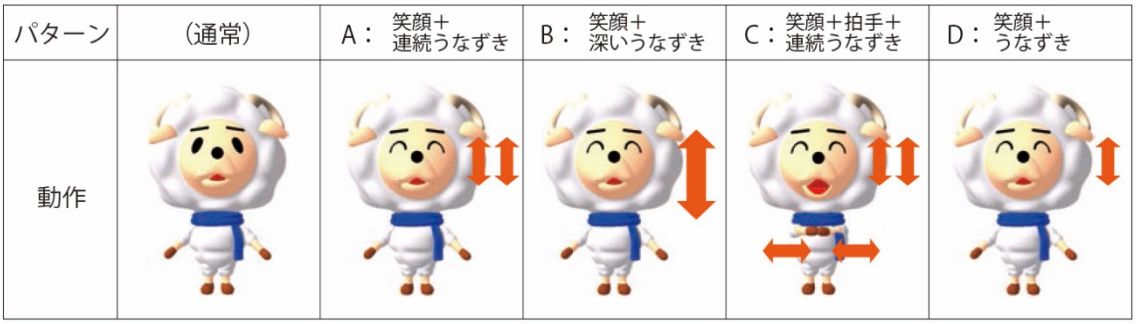

図 2.7　強い同調動作

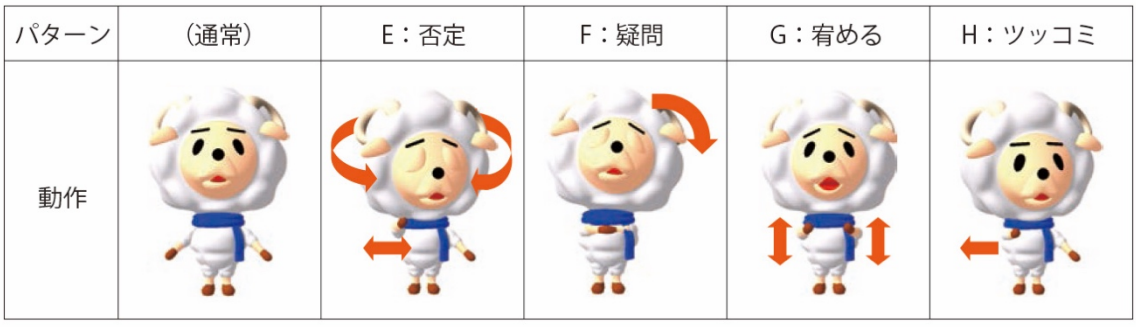

図 2.8　緩和動作

本システムにおけるキャラクタが行う各動作の提示確率を検討するために, 同調動作, 緩和動作それぞれについて一対比較実験を行った．各 4 パターンの動作をキャラクタシステムに付加したものを用い, 実験参加者に対し, 各 4 パターンの動作の中から 2 つ選び, 実際にシステムを使用させた上で, それぞれ強い同調動作・緩和動作として好ましいものを選択させた．発話内容として, 強い同調動作の際は過去の肯定的な出来事, 緩和動作の際は過去の否定的な出来事を思い出しながら 1 分間自由に発話させた．これを 1 人に対して各 6 通り（${}_4C_2$）行った．モードの提示順序はカウンターバランスをとり, 評価への影響を排除するために, 総当たりにより提示順序を変更した．実験参加者は 22～24 歳の学生 10 名（男性 8 名，女性 2 名）である．否定的な発話内容が聴取あるいは記録されることの影響を排除するため, 実験の様子は記録しなかった．

2.7.2 実験結果とシステム適用

各一対比較の結果を表 2.2，2.3 に示す．一対比較による評価を一義的に定めるために，Bradley-Terry モデル（$P_{ij} = \pi_i / (\pi_i + \pi_j)$，$\Sigma \pi_i = const.$ (= 100)，$\pi_i : i$ の強さ，$P_{ij} : i$ が j に勝つ確率）を想定した．π_i は 4 種類のモードの強さ（4 モードの合計 100）を表し，このモデルを想定することにより，一対比較に基づく好ましさを一義的に定めることができる．モードの強さ π を最尤推定した結果を図 2.9，2.10 に示す．強い同調動作は，A の連続うなずき＋笑顔の動作，C のうなずき＋拍手＋笑顔（笑い）の動作，B の深いうなずき＋笑顔の動作，D のうなずき＋笑顔の動作の順で好まれる結果となった．また，緩和動作は，G の口を開閉＋腕を上下（宥める）動作，E の手と頭を左右に振る（否定）動作，H の手を横に出す（ツッコミ）動作，F の腕を組む＋首をかしげる（疑問）動作の順で好まれる結果となった．

本システムではこの実験結果から得られた各動作の好みの強さ π を確率分布とし，それぞれの強い同調動作・緩和動作の選択確率とした．この選択確率に従って，身体的引き込み反応に加えて，感情極性推定から算出した感情極性値が予め設定した Positive の閾値を上回った，あるいは Negative の閾値を下回った場合に，キャラクタが強い同調動作，あるいは緩和動作を行うキャラクタシステムを開発した．開発した

システムにおいて, 音声入力が終了してから音声が認識された結果が表示されるまでの時間は，動画による評価で約 167 ms 程度であり，またその時点からキャラクタ動作が開始されるまでに 100 ms 程度で，合わせて約 267 ms 程度の時間遅れが生じている．従来の InterActor は二値化した音声入力について約 133 ms のハングオーバー処理を行っているため，従来のシステムとの時間遅れの差は約 133 ms 程度となる．先行研究によって CG アバタを介したコミュニケーションにおいて, 音声通話における遅延については 400 ms までが推奨され［2.7］，また 500 ms 程度の通信遅延では評価に影響を与えないことが明らかになっており［2.8］, 本システムの時間遅れについて大きな影響はないと考えられる．

表 2.2　強い同調動作の一対比較結果

	A	B	C	D	計
A		6	5	8	19
B	4		4	5	13
C	5	6		7	18
D	2	5	3		10

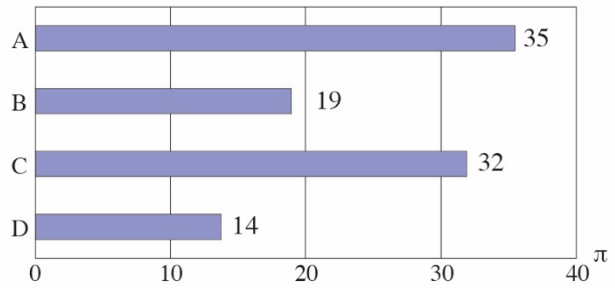

図 2.9　強い同調動作の各動作に対する好みの強さ π

表 2.3　緩和動作の一対比較結果

	E	F	G	H	計
E		7	3	9	19
F	3		2	4	9
G	7	8		6	21
H	1	6	4		11

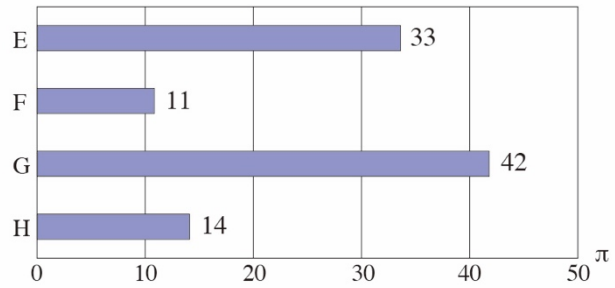

図 2.10　緩和動作の各動作に対する好みの強さ π

2.8 おわりに

本章では，従来の音声駆動型身体的引き込みキャラクタ InterActor に音声認識を導入して発話内単語の感情極性から使用者の感情を推定し, 結果に基づいて反応動作を変化させる音声駆動型身体的引き込みキャラクタシステムの開発を行った. 認識した単語から感情極性値を算出し, Positive の閾値を上回った場合には強く同調する動作, Negative の閾値を下回った場合にはそれを打消し, 緩和させる動作を行う. 同調動作・緩和動作は 4 パターンずつ用意し, 実験から得られた各動作の好みの強さを確率分布とし，それぞれの強い同調動作・緩和動作の提示確率とした.

参考文献

［2.1］ 檀原龍正, 渡辺富夫, 大久保雅史, 音声駆動型身体引き込みキャラクタ InterActor が発話音声に与える効果, 日本機械学会論文集 C 編, Vol.71, No.712 (2005), pp.152–159.

［2.2］ 渡辺富夫, 人間－機械間の音声対話のための時間率適応化システム, 計測自動制御学会論文集，Vol.26, No.8 (1990), pp.50–55.

［2.3］ 高村大也，乾孝司，奥村学，スピンモデルによる単語の感情極性抽出，情報処理学会論文誌，Vol.47, No.2 (2006), pp.627–637.

［2.4］ Ekman, P., Friesen, W. V. and Ellsworth, P., What emotion categories or dimensions can observers judge from facial behavior emotion in the human face, New York: Cambridge University Press (1982), pp.39–55.

［2.5］ Parrott, W., Emotions in social psychology, Psycology Press (2001).

［2.6］ 李晃伸，河原達也，Julius を用いた音声認識インタフェースの作成，ヒューマンインタフェース学会誌，Vol.11, No.1 (2009), pp.31–38.

［2.7］ 伊藤憲三, 北脇信彦, 会話音声の時間的特徴量に着目した遅延品質評価法, 日本音響学会誌，Vol.43, No.11 (1987), pp.851–857.

［2.8］　石井裕，瀬島吉裕，渡辺富夫，通信遅延環境における自己の身体的アバタ動作遅延提示の効果，ヒューマンインタフェース学会論文誌，Vol.13, No.1 (2010), pp.23–30.

第 3 章

シナリオに基づく語りかけと二者対話による反応動作評価実験

3.1 はじめに

本章では，2 章で開発した単語感情極性に基づく音声駆動型身体的引き込みキャラクタシステムの評価実験を行った．まず，Negative または Positive と判定されたシナリオに基づくキャラクタへの語りかけによる自動応答エージェントシステムとしての評価実験を行った. 次に二者対話によるコミュニケーションインタフェースシステムとしての評価実験を行い，システムの有効性を示した.

シナリオに基づくキャラクタへの語りかけによる評価実験では, 正負の感情に対して適切な動作表現ができているかという観点から, あらかじめ用意したシナリオに基づいて, キャラクタへの語りかけによって評価した．とくに Negative なシナリオについては，本研究で課題とした使用者の負の感情を助長しないという観点から，負の感情を抑制する動作表現に加え，比較対象として，あえてさらに同調を促す動作表現を用いてシステムを評価している.

二者対話による評価実験ではキャラクタを対話者それぞれのアバタとして扱い, 二者間コミュニケーションによってシステムを評価した. 音声認識による感情極性に対応した動作は，相手キャラクタの聞き手動作として反映される.

3.2 Negative なシナリオに基づく評価実験

3.2.1 実験方法

まず Negative な内容のシナリオに基づくキャラクタへの語りかけによる評価実験を行った．本実験では，次に示すキャラクタの動作が異なる N1～N3 の 3 つのモードを用意した．

N1：従来の InterActor が行ううなずきや身振りなどの聞き手としての動作を行う
N2：N1 モードの聞き手としての動作に加え，本研究で開発した提案手法による感情推定を行い，推定結果が Negative の閾値を下回った場合にキャラクタが負の感情に対し強い同調動作を行う
N3：N1 モードの聞き手としての動作に加え，本研究で開発した提案手法による感情推定を行い，推定結果が Negative の閾値を下回った場合にキャラクタが負の感情を緩和させる動作を行う

システムを十分理解した上で評価を行うために，最初に各モードの違いと操作方法を説明した．その後実際にシステムを試用させた．説明にあたっては，実験者が統制された説明を行うことができるように，実験に関する運用手順書を作成した．モードの切り替えや操作は，実験者がマイクとスピーカにより指示した．実験参加者にはアンケート用紙をあらかじめ配布し，評価項目ごとに記入させた．ただし，各モードを使って語りかけている途中は，実験に集中させるためアンケート用紙に触らないよう指示した．

実験は，最初に N1～N3 の 3 つのモードのうち 2 つを提示し，一対比較法により各モードの比較を行わせた．実験参加者には総合的に良い方を選択させた．提示場面を変更し，3 回（${}_3C_2$）繰り返した．モードの提示順序はカウンターバランスをとり，評価への影響を排除するために，総当たりにより提示順序を変更した．次に，各モードに対し身体的コミュニケーション支援の観点から定めた 6 項目について 7 段階官能評価（中立 0）を行った．6 項目は，「楽しさ」「好み」「対話しやすさ（話しやすさ）」「安心感」「和み」「システムを使用したいか」とした．実験参加者に提示す

るシナリオは年齢や家族構成を始め, 背景や場面設定が記されているものを使用した（図 3.1）．実験参加者は 19～24 歳の学生 24 名（男性 12 名，女性 12 名）で，実験参加者には 500 円の図書カードを謝礼として渡した．図 3.2 は実験の様子である．

21 歳　大学 3 年生
家族構成:父親 51 歳　母親 48 歳　兄 25 歳の 4 人暮らし

夏季休暇も終了し、前期の成績の結果が返却された。必修科目については無事取得出来ており問題なかったが、選択科目をいくつか落としてしまっていた。現状、進級に必要な単位のギリギリで履修登録を行っていたため、後期の履修可能分を全て取得出来なければ進級が危うく、留年の危機が迫っている。
バイトや部活などもあるため忙しく、かなりの過密スケジュールになると予想されるため, このままでは後期の単位取得も難しくなるのではないかと不安に駆られている。そのため、ここ最近は寝つきが悪く、良質な睡眠がとれていない。来年度には 4 年に進級し、卒業研究に加え就職活動を始めなければならないため、これから先のことについて考えていると不安が募り、気分が滅入っている。
現在は両親などと話し合いを行いつつ、今後のことについて考えている。昨晩は精神的不安などからあまり食欲もわかず、あまり眠れなかったため、気分が落ち込んだまま 1 日を過ごしている。

場面:日曜の午後 10 時、自宅の自分の部屋。来週の月曜から後期の講義が開始となる。食後, 何をする様子もなく机に向かっている。

図 3.1　ネガティブなシナリオの例

図 3.2　実験の様子

3.2.2 実験結果

一対比較結果を表 3.1 に示す．一対比較による評価を一義的に定めるために Bradley-Terry モデルを想定し，強さ π を最尤推定した結果を図 3.3 に示す．N3 の「聞き手としての動作，および負の感情を緩和させる動作」のモードが N1 に対して 12.5 倍，N2 に対して 17.6 倍と突出して高く評価されており，続いて N1 の「うなずきや身振り手振りなどの聞き手動作」のモード，N2 の「聞き手としての動作，および負の感情に対する強い同調動作」のモードの順に評価されている．

表 3.1　一対比較結果

	N1	N2	N3	計
N1		14	1	15
N2	10		2	12
N3	23	22		45

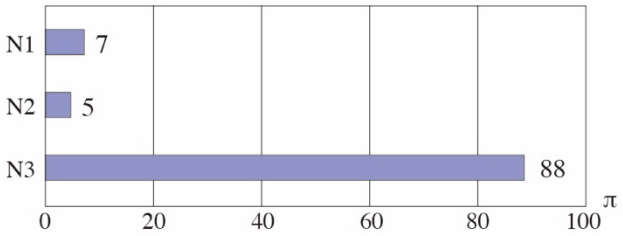

図 3.3　各モードの強さ π

次に，7 段階官能評価の結果について各項目の平均及び標準偏差を図 3.4 に示す．3 条件間の比較として Friedman 検定を実施した結果，すべての項目において有意水準 5 %で有意差が認められた．さらに多重比較として Wilcoxon の符号順位検定を行った結果，N3 モードと N1 モードの間では「安心感」の項目において有意水準 0.1 %で有意差が認められ，「楽しさ」「好み」「和み」の項目において有意水準 1 %で有意差が認められ，「対話しやすさ（話しやすさ）」「システムを使用したいか」の項目において有意水準 5 %で有意差が認められた．また，N1 モードと N2 モードの間では，「好み」の項目において有意水準 0.1 %で有意差が認められ，「システムを使用したいか」の項目において有意水準 1 %で有意差が認められ，「対話しやすさ（話しやすさ）」「安心感」の項目において有意水準 5 %で有意差が認められた．さらに，N3 モードと N2 モードの間では，「好み」「対話しやすさ（話しやすさ）」「安心感」

「システムを使用したいか」の項目において有意水準 0.1 %で有意差が認められ，「和み」の項目において有意水準 1 %で有意差が認められた．

実験時に得られた自由記述回答を表 3.2 に示す．

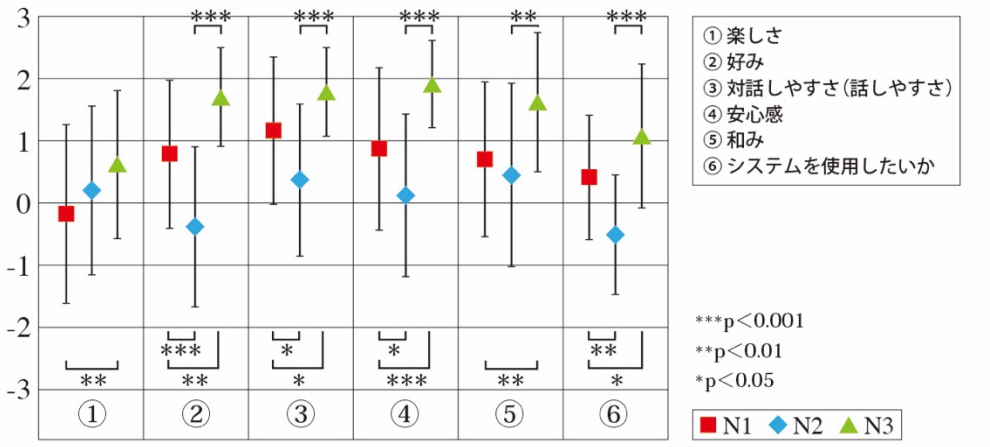

図 3.4　7 段階官能評価結果

表 3.2　自由記述

肯定的な意見
・N3 モードの方が対話しているような感じがするので良かった ・N3 は辛いときにキャラクタが「そうじゃない」「大丈夫」と動作してくれたのがよかった ・悩んでいる場面等では盛り上げてもらうよりも聞いてもらうだけの方が安心できた ・N1 はうなずくだけだったけど話をしっかりと聞いてくれる感じ ・N3 は「そんなことないよ」と言われている感じでいちばん好き ・どのタイプも会話にうなずいてくれるのは話しやすくてよかった ・N3 モードは勇気づけてくれている感じがして元気になれそうだと思った
否定的な意見
・不安なことを話しかけたが，N2 は笑顔が返ってきたので違和感があった ・N2 は本当につらいときに笑顔というか笑っていたらちょっと嫌だった ・N2 はつらい気持ちなのに，それを打ち消すように笑ったりしていて，元気になれる時もあるかもしれないけど，自分の気持ちを否定されているように感じた

3.2.3 考察

実験の結果，一対比較では N3 の「聞き手としての動作，および負の感情を緩和させる動作」のモードが突出して高く評価された．また，7 段階官能評価の結果において，N1 モードに対し，提案手法を用いた N3 モードで「楽しさ」「好み」「安心感」「和み」の項目で有意差が確認された．また N2 モードに対し，提案手法を用いた N3 モードで，「好み」「対話しやすさ（話しやすさ）」「安心感」「和み」「システムを使用したいか」の項目で有意差が確認された．また，「聞き手としての動作，および負の感情を緩和させる動作」の N3 モードに関する自由記述では「辛いときにキャラクタが「そうじゃない」「大丈夫」と動作してくれたのがよかった」や「勇気づけてくれている感じがして元気になれそうだと思った」といったコメントがあり，N3 モードが好意的に受け入れられていることが確認された．一方で N2 モードでは「N2 は本当につらいときに笑顔というか笑っていたらちょっと嫌だった」や「N2 はつらい気持ちなのに，それを打ち消すように笑ったりしていて，元気になれる時もあるかもしれないけど，自分の気持ちを否定されているように感じた」などといったコメントが見られた．これらのことから，Negative な内容の発話を行った際に，それを否定，打ち消すことで発話者に対し和みなどを与えることができ，負の感情で発話された内容に対する反応として，好ましいと判断されたことがわかる．さらに，7 段階官能評価において N2 モードが N1 モードに対し低く評価されていることから，使用者の負の感情に対して適切な同調・緩和動作を用いることが必要であるといえる．

3.3 Positive なシナリオに基づく評価実験

3.3.1 実験方法

次に Positive な内容のシナリオ（図 3.5）に基づくキャラクタへの語りかけによる評価実験を行った．実験環境および実験者からの指示等は前節と同様である．前節の実験結果より，負の感情に対しては適切な同調・緩和動作を用いることが必要であることが確認された．本実験では，正の感情に対する強い同調動作の効果を検討するために，次に示す P1，P2 の 2 つのモードを用意した．

P1：従来の InterActor が行ううなずきや身振りなどの聞き手としての動作を行う

P2：P1 モードの聞き手としての動作に加え，本研究で開発した提案手法による感情推定を行い，推定結果が Positive の閾値を上回った場合に，キャラクタが正の感情に対し強い同調動作を行う

実験はまず，P1，P2 の 2 つのモードを用いて各モード 1 分間ずつ使用させた後，一対比較によって総合的に良かった方を選択させた．次に，各モードに対し身体的コミュニケーション支援の観点から定めた前節と同じ 6 項目について 7 段階官能評価（中立 0）を行った．実験参加者は 20～24 歳の学生 24 名（男性 12 名，女性 12 名）で，実験参加者には 500 円の図書カードを謝礼として渡した．

21 歳　大学 3 年生
家族構成：父親 51 歳　母親 48 歳　兄 25 歳の 4 人暮らし

私の誕生日パーティーが開かれた。友人たちがサプライズで企画してくれていたようだった。
いつも通り一日の講義を受け終え自宅に帰ろうとすると、友人からある場所に来てほしいとの連絡があった。何も知らされていなかったため疑問に思いつつ一人で指定された部屋に入ると、綺麗に装飾された部屋に食事などがたくさん用意されており、友人たちが私を迎え入れてくれた。
驚きとともに、嬉しさがこみ上げてきた。おいしい食事をたくさん食べることが出来たり、友人たちが余興を行ってくれたり、さらにはプレゼントまで用意してくれるなど、申し訳ないほどもてなしを受けた。非常に楽しい会であったため、これ以上ない幸せな気分である。
最近は学業やバイト、部活などで忙しく、みんなで集まって遊ぶことなどが出来なかった。そのため、今回の誕生日パーティーではみんなが揃い有意義な時間を過ごすことで心身ともにリフレッシュすることが出来た。
現在は片付けを終え、みんなからもらったプレゼントを開封しながら誕生日パーティーの余韻に浸っている。そのため、気分が高まった状態で過ごしている。

場面：誕生日パーティーの片付けを終え自宅に帰宅。自分の部屋でのんびりとしている。

図 3.5　ポジティブなシナリオの例

3.3.2 実験結果

一対比較の結果，実験参加者 24 名全員が P2 の「聞き手としての動作，および正の感情に対する強い同調動作」のモードを選択した．また 7 段階官能評価の結果について各項目の平均及び標準偏差を図 3.6 に示す．Wilcoxon の符号順位検定を行った結果，P2 モードと P1 モードの間で全ての項目において有意水準 0.1 %で有意差が認められた．

実験時に得られた自由記述回答を表 3.3 に示す．

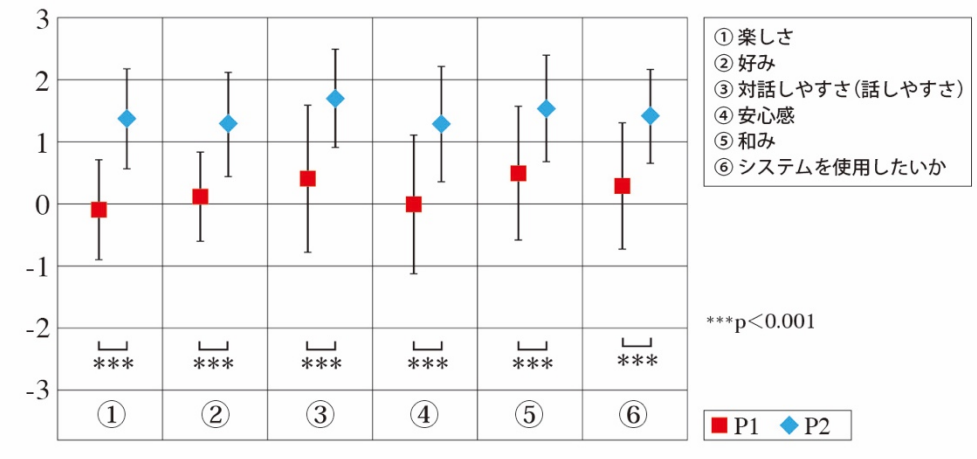

図 3.6　7 段階官能評価結果

表 3.3　自由記述

肯定的な意見
・P2 は自分が話したことに対して反応があったのでおもしろかった ・羊がニコニコしてくれるとこちらも楽しくなって話したいなと思った ・P2 の方が話しやすかった ・笑ってくれたりした方が反応を感じられてこちらも話しやすかった ・P1 はうなずくだけだったけど話をしっかりと聞いてくれる感じ ・忙しい話をしているときは P1 モードもいいと思った
否定的な意見
・P1 モードは悩みや真剣に話を聞いてほしいときには話しやすいが楽しいときには話しづらかった

3.3.3 考察

P2 モードは一対比較，7 段階官能評価ともに高く評価された．また P2 モードに関する自由記述では「羊がニコニコしてくれるとこちらも楽しくなって話したいなと思った」や「笑ってくれたりした方が反応を感じられてこちらも話しやすかった」といったコメントがあり，P2 モードが好意的に受け入れられていることが確認された．また， P1 モードに関する自由記述では「P1 モードは悩みや真剣に話を聞いてほしいときには話しやすいが楽しいときには話しづらかった」や「忙しい話をしているときは P1 モードもいいと思った」など，状況に応じたキャラクタの動作提示が必要であるということが示唆された．これらの結果より，Positive な内容の発話を行った際に，それに同調し盛り上げることで，発話者に対して楽しさ，話しやすさなどを与えることができ，正の感情を盛り上げるための好ましい効果が得られたと考えられる．

3.4 二者対話によるシステム評価実験

3.4.1 実験方法

前章の実験は，発話を Negative または Positive な内容に限定した上でのキャラクタに対する語りかけによる実験であった．実際にシステムを使用した際の評価を行うためには，一般的な対話状態における発話者の感情推定の結果に応じて，同調・緩和動作を行うキャラクタシステムを用いたコミュニケーション実験を行う必要がある．そこでキャラクタシステムに通信機能を導入することにより，遠隔で二者対話を実現するコミュニケーションシステムを開発した（図 3.7，3.8）．2.6 節で構築したシステム構成を用いて，2 台の PC を 1 Gbps Ethernet で接続し，音声を送受信している．

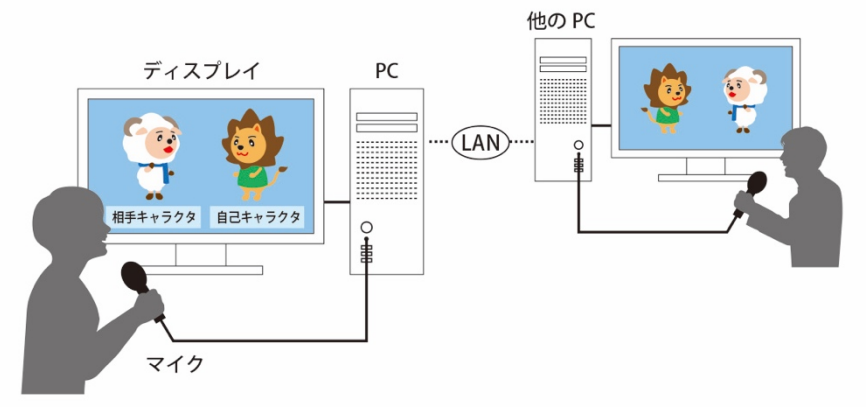

図 3.7　対話システムの概略図

図 3.8　対話システムの使用図

図 3.9　実験の様子

システムの評価として，二者対話による評価実験を行った．実験の様子を図 3.9 に示す．本実験では，次に示すキャラクタの動作が異なる α，β の 2 つのモードを用意した．

α：発話時の身振り手振りなど話し手動作に加え，うなずきなどの聞き手としての動作を行う

β：α モードの従来動作に加え，本研究で開発した提案手法による感情推定を行うもので，推定結果が Positive の閾値を上回った場合にキャラクタが正の感情に対し強い同調動作を行い，推定結果が Negative の閾値を下回った場合にキャラクタが負の感情を緩和させる動作を行う

実験参加者は 19～22 歳の男女学生で，同性同士の友人関係である者を 2 人 1 組とした 12 組 24 名（男性 6 組，女性 6 組）である．実験参加者には 500 円の図書カードを謝礼として渡した．実験は 2 つの部屋を用い，それぞれ別の個室に分かれてシステムを用いて日常の談話をさせた．最初に，各モードの違いと操作方法を説明した後，実際にシステムを数分間試用させた上で実験を行った．

実験はまず，α，β の 2 つのモードを各 3 分間ずつ使用させた後，一対比較によって総合的に良かった方を選択させた．次に，各モードに対し身体的コミュニケーション支援の観点から定めた前節と同じ 6 項目について 7 段階官能評価（中立 0）を行った．

3.4.2 実験結果

一対比較の結果を図 3.10 に示す．β モードの「聞き手としての動作，および正の感情に対する強い同調動作，および負の感情を緩和させる動作」のモードが高く評価されていることが分かる．次に，7 段階官能評価について各項目の平均及び標準偏差を図 3.11 に示す．Wilcoxon の符号順位検定を行った結果，β モードと α モードの間で「楽しさ」の項目において有意水準 0.1 %で有意差が認められ，「好み」「和み」「システムを使用したいか」の項目において有意水準 1 %で有意差が認められ，また「対

話しやすさ（話しやすさ）」「安心感」の項目では有意水準 5 %で有意差が認められた.

実験時に得られた自由記述回答を表 3.4 に示す.

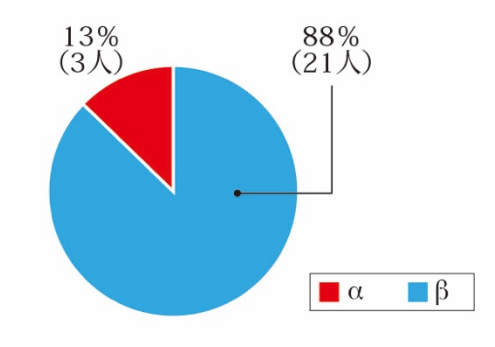

図 3.10　一対比較結果

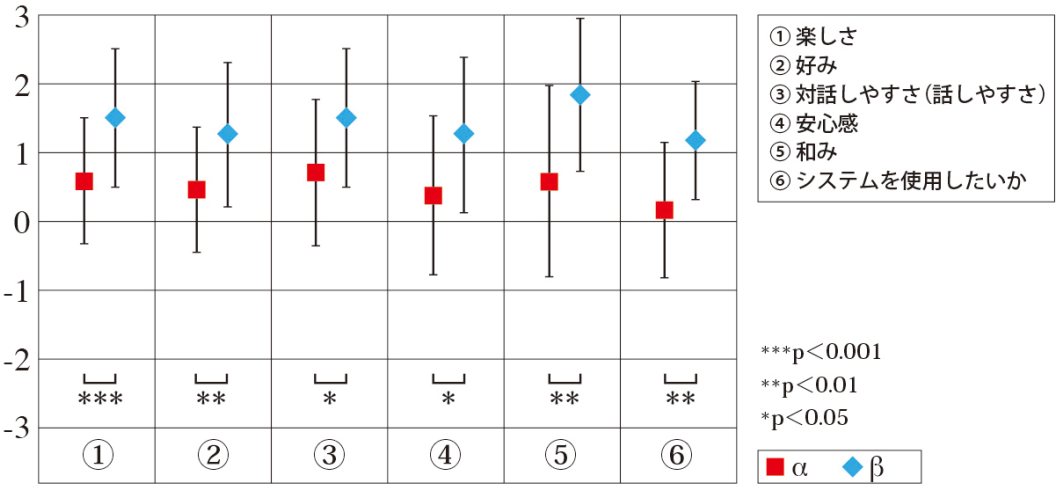

図 3.11　7 段階官能評価結果

表 3.4　自由記述

肯定的な意見
・βモードで苦労話をしたときに緩和動作をとってくれてとても和んだ ・少し違うワードに対しての動作もあったが，目の形が変化したりして話しやすい雰囲気になった ・ネガティブな発言に対し，「まぁまぁ」のような動作があったのはよかった ・キャラが同調してくれると相手により伝わりやすくいいと思った ・「辛い」「楽しい」という言葉に対し，相手のキャラクタが上手く反応していた ・うなずく動作や笑う動作がある方がリアルに話しているように見えて話がしやすく思えた ・相手の顔が直接見えなくても，キャラクタが反応してくれるので話しやすかった
否定的な意見
・自分の声とキャラの動きが一致しているときもあれば，たまに遅れて測定しきれない部分もあったので判断が難しかった ・Negative なワードを言ってない時に動作をしてくれることが何回かあったのでそっちに目が行ってしまい，話の内容が頭に入ってこないことがあった ・「やばい」のようなマイナスでもプラスでもとらえることのできる言葉にも対応できるとよりよくなると思った

3.4.3 考察

実験の結果，β モードが一対比較，7 段階官能評価ともに高く評価されており，システムの有効性が示された．β モードに対する自由記述では「うなずく動作や笑う動作がある方がリアルに話しているように見えて話がしやすく思えた」や「苦労話をしたときに緩和動作をとってくれてとても和んだ」といったコメントがあり，β モードが好意的に受け入れられていることが確認された．一方で「自分の声とキャラの動きが一致している時もあれば，たまに遅れて測定しきれない部分もあったので判断が難しかった」「Negative なワードを言ってない時に動作をしてくれることが何回かあったのでそっちに目が行ってしまい，話の内容が頭に入ってこないことがあった」などといったコメントもあり，発話文取得の際の認識精度や，動作の提示タイミングなどに対して改善すべき意見が得られた．

3.5 おわりに

本章では 2 章で開発した，音声駆動型身体的引き込みキャラクタ InterActor に音声認識を導入し，発話感情推定を行うキャラクタシステムの評価実験を行った．

Negative または Positive と判定されたシナリオに基づくキャラクタへの語りかけによる自動応答エージェントシステムとしての評価実験では, 感情極性に対応して適切に反応動作を行うことが評価された．また，二者対話によるコミュニケーションインタフェースシステムとしての評価実験により，システムの有効性が示された．

第 4 章

発話活性度及び感情極性に基づく反応動作生成システム

4.1 はじめに

2 章及び 3 章で，単語感情極性から話者の感情を推定し，結果に基づいて反応動作を変化させることで，話者の否定的な発話を抑制，または肯定的な状態を保持することができる身体的引き込みキャラクタシステムを開発し，システムの有効性を示した．しかし実際の人間の感情には多くの感情区分が存在する［4.1，4.2］．単語の感情極性対応では，その単語が良い印象を持つか（Positive），悪い印象を持つか（Negative）の判断しか行うことができず，使用者の多様な感情に対し，それぞれの場合に適切な支援を行うことが困難であると推測される．

そこで本章では，人の様々な感情を Arousal（覚醒度）と Valence（快・不快）の二軸で分類するラッセルの円環モデル［4.3］に基づき，使用者の発話活性度を Arousal，使用者の発話内容（Positive/ Negative）を Valence として位置付けた状態推定モデルを定義し，結果に基づいて反応動作を行う身体的引き込みキャラクタシステムを開発した．また開発したシステムを用いて，発話活性度を考慮した状態推定の有効性を検討する評価実験を行った．

4.2 コンセプト

人の感情を Positive な活性状態では興奮や喜び，Negative な活性状態では怒りや心配などの感情に分類したラッセルの円環モデルが提案されている（図 4.1）［4.3］．本システムでは発話時間率 Speech Activity（SA）を発話活性度とし，ラッセルの円環モデルに基づいて, Arousal を SA, Valence の Pleasant を Positive, Unpleasant を Negative とすることで 5 つの状態を推定したモデルを定義した（図 4.2）．キャラクタは話者の語りかけに対して自動生成される身体的引き込み動作に加えて, 推定した 5 つの状態に対応した動作を行う．システムのコンセプトを図 4.3 に示す.

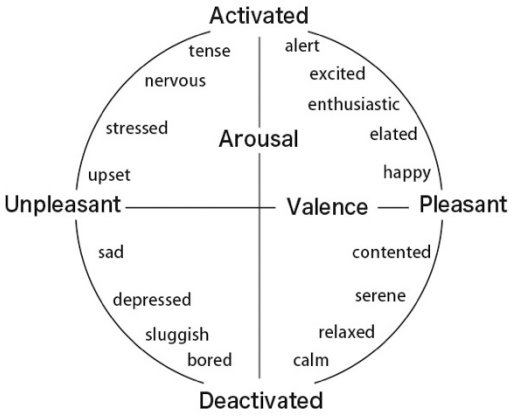

図 4.1　ラッセルの円環モデル

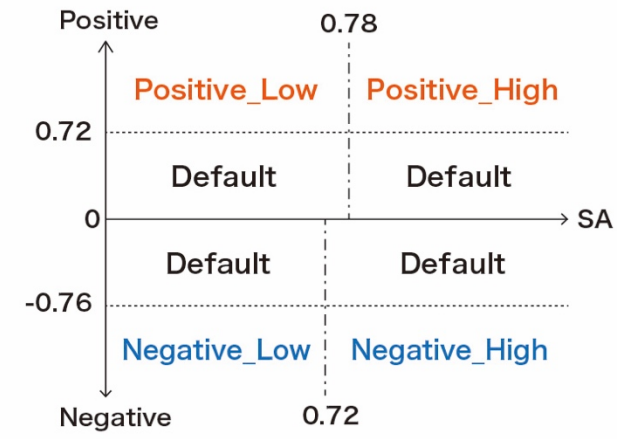

図 4.2　状態推定モデル

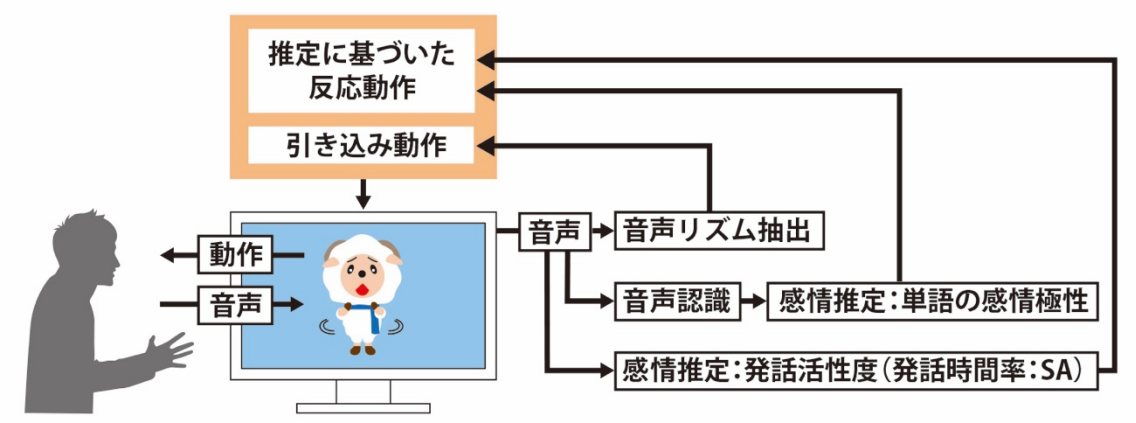

図 4.3　コンセプト

4.3 システム構築

4.3.1 閾値の決定

Positive，Negative の閾値は，前章の発話内単語の感情極性を用いたシステムと同様とした．入力音声を RealSence SDK（Intel）の音声認識ツールによってテキスト化し，得られたテキストを感情極性対応表内の単語と照合し，一致する単語があればその感情極性値を抽出する．　SA の閾値は，6 名の実験参加者に「楽しかった話」「悲しかった話」をキャラクタに 2 分間程度話させ，平均した値を SA の閾値（Positive: 0.78，Negative: 0.72）とした．

4.3.2 発話活性度算出方法

図 4.4 のように，2 値音声をハングオーバー（133 ms）処理した i ユニット（1 ユニットは ON 開始時点から次の ON 開始時点までの区間）での ON 区間を $T(i)$，OFF 区間を $S(i)$とし，式（4.1）を用いて逐次 SA を算出した．

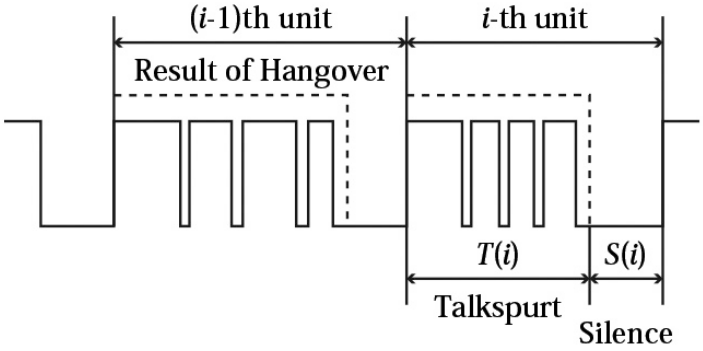

図 4.4　Burst-pause of speech with hangover.

$$SA(i) = \frac{\sum_{j=1}^{6} T(i-j)}{\sum_{j=1}^{6}\left(T(i-j) + S(i-j)\right)} \tag{4.1}$$

4.3.3 動作表現

感情極性値及び SA によって推定された状態に対応した動作として，それぞれの状態で 2 パターン用意した．Positive_Low では状態を維持するために 1：うなずき＋笑顔，2：深いうなずき＋笑顔の同調動作，Positive_High では盛り上がりに呼応するために 3：連続うなずき＋笑顔，4：うなずき＋拍手＋笑顔の盛り上げ動作を行う．Negative_Low では肯定的な発話へ変化させるために 5：手と頭を左右に振る，6：体を傾けて手を横に出す（ツッコミ動作）の否定動作，Negative_High では落ち着かせるために 7：口の開閉＋体を傾ける＋腕を上下させる（宥める），8：腕を組む＋首をかしげる（疑問）の緩和動作を行う（図 4.5）．また，感情極性値が Positive の閾値よりも小さく，Negative の閾値よりも大きい場合（Default）は通常のうなずき動作を行う．

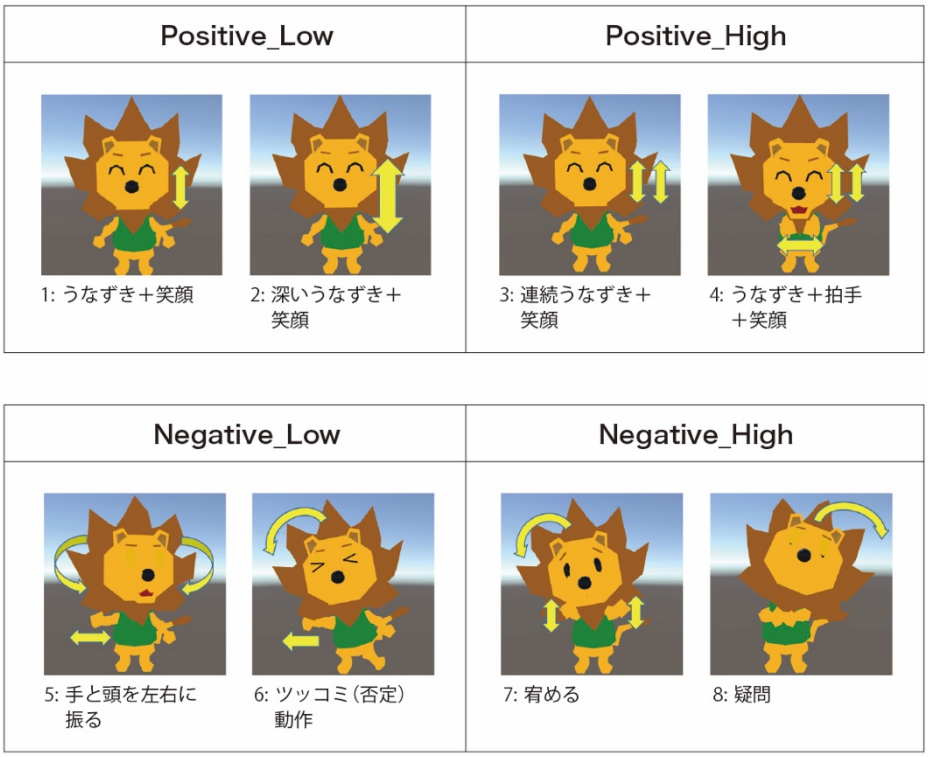

図 4.5　動作表現

4.4 発話活性度及び感情極性に基づく反応動作生成システムの評価実験

4.4.1 実験方法

開発したシステムの有効性を評価するために，実験参加者の感情に応じた内容について自由に発話させる実験を行った．実験は次に示す 2 つのモードを用意した．

A：発話内単語の感情極性を用いて推定した 3 つの状態（Positive/Negative/Default）に対応してキャラクタが動作を行う．反応動作は図 2.7，2.8 の動作パターンと同様である

B：SA および発話内単語の感情極性を用いて推定した 5 つの状態（Positive_Low/ Positive_High/Negative_Low/ Negative_High/Default）に対応してキャラクタが動作を行う

実験参加者には発話感情（正または負）を指示し，自由に発話させた．指示した発話感情の順序は順序効果を考慮して，総当たりによるカウンターバランスをとった．

まず，各モード 3 分間ずつ話をさせ，どちらが良かったかを一対比較法により比較させた．次に，モード毎に 3 分間話をさせ，終了時に「楽しさ」「好み」「話しやすさ」「盛り上がり（Positive）／緩和されたか（Negative）」「親近感」「システムを使用したいか」の 6 項目について 7 段階官能評価を行った．また，自由記述欄に気づいたことを記入させた．実験参加者は 18〜23 歳の学生 20 名（男性 10 名，女性 10 名）である．実験参加者には 500 円の図書カードを謝礼として渡した．

4.4.2 実験結果

Positive な発話内容による実験について，一対比較の結果，両モードに大きな差は見られなかった（表 4.1）．また，7 段階官能評価の結果についても大きな差は見られず（図 4.6），Wilcoxon の符号順位検定を行ったが有意差は認められなかった．Negative な発話内容による実験についても，一対比較の結果，両モードに大きな差は見られなかった（表 4.2）．また， 7 段階官能評価の結果についても大きな差は見られず（図 4.7），Wilcoxon の符号順位検定を行ったが有意差は認められなかった．

実験時に得られた自由記述回答を表 4.3，4.4 に示す．

表 4.1　一対比較結果：Positive

	A	B	計
A		12	12
B	8		8

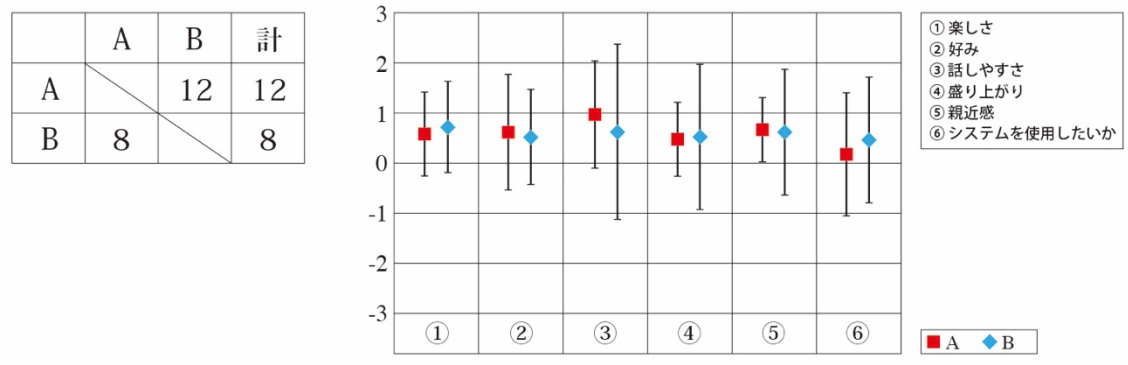

図 4.6　7 段階官能評価結果：Positive

表 4.2　一対比較結果：Negative

	A	B	計
A		11	11
B	9		9

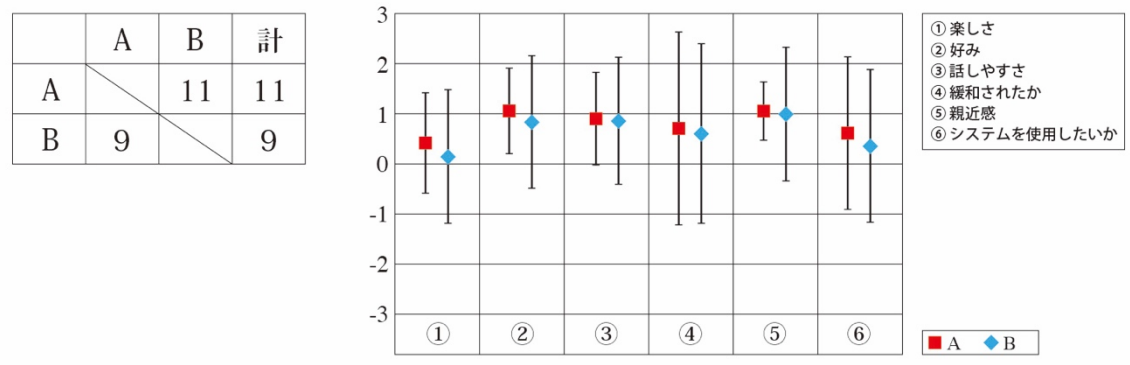

図 4.7　7 段階官能評価結果：Negative

表 4.3　自由記述：Positive

肯定的な意見
・話の頂点で笑ってくれたときはうれしかった ・手をたたく動作はうれしかった ・要所要所でリアクションをしてくれて話しやすかった
否定的な意見
・モードの違いがわかりにくい ・ライオンのリアクションが小さい

表 4.4　自由記述：Negative

肯定的な意見
・辛いことを話したときに慰めてくれる感じがしたのでうれしかった ・B モードはすごく聞いてくれている感じがした
否定的な意見
・目を閉じて首を横に振る動作（ツッコミ）を入れるのは話を聞いてほしいので良くないと思う ・A モードで時々入るツッコミが気になる ・愚痴を話すときは派手なリアクションをとられるよりも，ゆっくり頷いてくれた方が話しやすかった

4.4.3 考察

両モードに大きな差が見られなかった原因として, モードの判別の難しさが考えられる. 本実験では自由に発話させたが, 実験参加者に対してシステムを使用する際の感情を指示したものの, Positive/Negative の閾値を超える単語があまり発話されなかったために通常のうなずき動作が提示される場合があった. 閾値を超える単語が発話された場合においても, A モードは 2.7.2 項の実験結果から定めた各動作の提示確率によって出現する頻度の高い動作があり, B モードは SA が変化しない場合に同じ状態が続くことで出現しやすい動作があることがモードの判別の難しさに繋がっている. また, SA が変化しないことにより出現する動作の多様性が失われたことで評価に繋がらなかった可能性もある.

キャラクタの動作表現については, Positive な発話内容に対する動作については肯定的な意見が見られた. 一方で, Negative な発話内容に対する動作については, 宥める動作は評価が高く, 首を横に振るなどの否定的な動作については「ツッコミを入れるのは, 話を聞いてほしいのでよくない」などの意見が見られ, 評価が低かった. 各状態に適した反応動作の検討を行うことで, より話者の状態に対応したシステムになると考えられる.

4.5 おわりに

本章では, 発話時間率に基づく活性度および発話内単語の感情極性を用いて話者の状態を推定し, 推定に基づき反応動作を行う身体的引き込みキャラクタシステムを開発した. 前章までの発話内単語の感情極性による状態推定システムとの比較実験を行ったが, 評価実験における活性度の状態変化が小さいことでモードの判別がつきにくかった可能性もあり, 十分な効果は認められなかった.

参考文献

［4.1］ Ekman, P., Friesen, W. V. and Ellsworth, P., What emotion categories or dimensions can observers judge from facial behavior emotion in the human face, New York: Cambridge University Press (1982), pp.39–55.

［4.2］ Parrott, W., Emotions in social psychology, Psycology Press (2001).

［4.3］ Russell, J. A. “A circumplex model of affect”, Journal of Personality and Social Psychology, Vol.39, No.6 (1980), pp.1161–1178.

第 5 章

音声相槌を伴うシステム

5.1 はじめに

前章までの検討の結果, 単語の感情極性に基づく状態推定と反応動作生成については効果が認められたが, 判定を詳細にした発話活性度を考慮した状態推定では効果は認められなかった．今後，状態に応じた反応動作を検討し，種類を増やしていくことは可能であるが，サムズアップは日本や英語圏では「Good」を意味するなど，動作には言語的な意味との関連が深いものが多くある．そのため，反応動作を多様化した場合, 使用者の話の文脈によっては, 応答として違和感のある提示になる可能性がある．

発話特性に関して一般に人同士の対話では，肯定よりも否定の方が発話のタイミングが遅いという指摘がある［5.1，5.2］．日常生活の中で，相手の発話に対して同調的に身体動作で応答しつつも，少し遅らせて音声相槌を発することも見られ，うなずきに相当する音声相槌を意味的な解釈として出力させることで使用者の発話感情に対応したシステムとして活用できる可能性がある．しかし従来の InterActor は，発声の有無（ON-OFF）に基づいてうなずきなどの身体動作が生成されることによる効果は確認してきたが，音声による応答機能は実装されておらず，その効果も確認されてこなかった．

そこで本章では, 身体的引き込み技術の音声対話エージェントへの応用を視野に入れ，iRT によるうなずきの出力タイミングに基づく音声相槌出力について検討した．従来の InterActor に音声相槌を付加したキャラクタシステムを開発し，自動生成する音声相槌について, その出力タイミングとキャラクタ表示による効果および音声相槌の出力頻度に着目した評価実験を行った．頻度の評価実験では，日本語対話特性を考慮した頻度で音声相槌を出力するシステムを構築し, 効果を確認するとともにシステムの有効性を示した．

5.2 コンセプト

身体的リズムの引き込みを基盤として，そのリズム同調を損なうことなく，実際の人同士の対面コミュニケーションリズムに基づいた音声相槌を行い, より円滑なコミュニケーションを促す音声応答を伴う音声駆動型身体的引き込みシステムを提案する（図 5.1）．

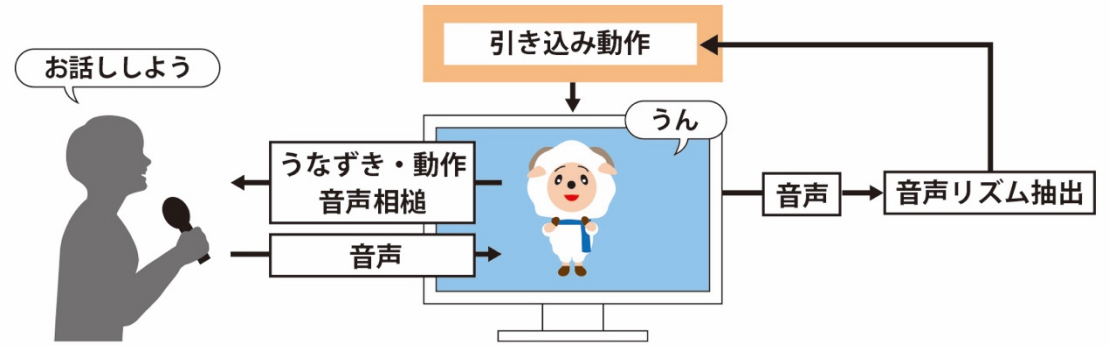

図 5.1　コンセプト

まず，使用者の語りかけに対して，その音声からキャラクタがうなずき動作を行うべきタイミングの推定を行う．音声相槌の場合，全てのうなずき動作に音声フィードバックを与えると使用者の発話を阻害する可能性があるため, 本システムではうなず

き動作の推定値より高い閾値を指定することで, うなずき動作に対する音声相槌の出現頻度を調整することができる. 人同士の対面コミュニケーションを擬似的に再現することで, 使用者への引き込み効果が高まり, 身体を介したコミュニケーションをさらに支援する効果があると期待される.

5.3 音声相槌の追加

本研究における音声相槌の生成タイミングの推定は，InterActor のインタラクションモデルで推定したマクロ層の予測値を用いる. 自動生成されたうなずき反応タイミング全てで音声相槌を行うのではなく, 頻度を調整するために, 別の閾値を設定する. 音声の ON-OFF に対し，2.4 節で述べた 133 ms でハングオーバー処理を施したデータから，図 5.2 におけるうなずき動作の有無を推定する閾値 Threshold より高い閾値 Threshold A を用い，それを超えた場合に，ミクロ層での閾値 Threshold 1 を超えた時点で，キャラクタのうなずき動作に対し音声相槌を付加することで，音声相槌の出現頻度を下げることができる．音声相槌に使用する発話として，感嘆詞「うん」を使用する. 感嘆詞「うん」は主に親しい関係で発せられ，多様な意味，機能を持つ［5.3］. 短い下降調の「うん」は比較的カジュアルな状況において出現するが［5.4］，本研究では発話促進の観点から「「聞き役中」の聞き手の反応」［5.5］として用いる．再生する音声の長さは 220 ms 程度で，短い下降調で基本周波数は 440 Hz であり，うなずきの動作は 500 ms 程度である.

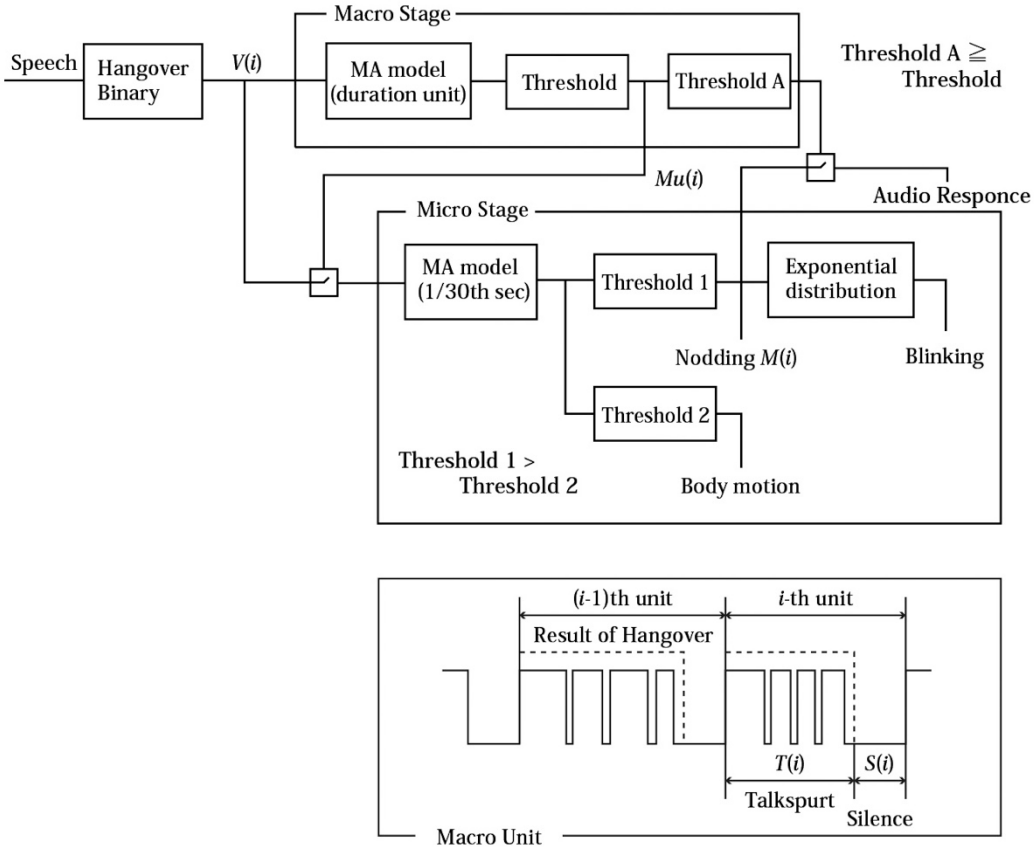

図 5.2　音声相槌を伴う InterActor のインタラクションモデル（聞き手）

5.4 音声相槌を伴う InterActor の評価実験

5.4.1 実験方法

iRT によるうなずきの出力タイミング予測を用いた音声相槌提示手法の効果を検討するために評価実験を行った．既述の通り，音声相槌の自動生成については先行研究でも検討されたことがなく，その効果を確認する必要がある．そのため本実験では，音声相槌のみによる検討として，キャラクタ表示しない状態での比較として 3 モード用意した．A モードは音声相槌無し，B モードはうなずき反応モデルによる音声相槌，C モードは B モードから 300 ms 遅延させた音声相槌である．この音声相槌の出

力タイミングについては，先行研究［5.6］において身体動作に対し音声を 300 ms 遅らせた場合がもっとも高く評価されていることから，音声相槌はうなずき動作よりも 300 ms 遅らせて提示した．またキャラクタ表示した状態での比較として，音声相槌無しの α モード，B，C モードに対応した音声相槌のある β，γ モードの 3 モードを加えた計 6 モードで評価を行った．

モードの提示順は，キャラクタ表示の有無で各 1 セットとし，A B C および α β γ の各 3 モードでカウンターバランスをとった．また順序効果を考慮し，キャラクタ表示の前後を入れ替えた．実験者が 3 つのモードから 2 つのモードを抽出し，実験参加者に 1 つのモードに対し自由なテーマで 1 分間の語りかけを行わせた．どちらが総合的に良かったかを聞く一対比較を行い，これを 3 通り（${}_3C_2$）行わせた．その後，改めて各 3 モードを使用させ，1 つのモードに対し 2 分間の語りかけを行わせ，5 項目（好み，楽しさ，話しやすさ，親近感，システムを使用したいか）について 7 段階官能評価（中立 0）を行わせた．実験参加者は 18〜24 歳の男女各 12 名の計 24 名で，実験参加者には 500 円の図書カードを謝礼として渡した．対話の様子は，PC 画面と，各対話者を後方から撮影した映像（図 5.3）を分割器で 2 分割した画像を生成し，録画した．

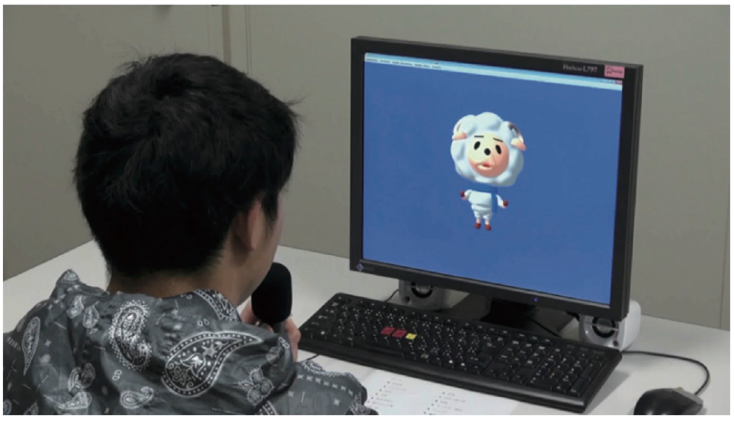

図 5.3　実験の様子

5.4.2 実験結果

一対比較の結果を表 5.1 に示す．一対比較による評価を一義的に定めるために，Bradley-Terry モデルを想定し，モードの強さ π を最尤推定した結果を図 5.4 に示す．その結果，キャラクタ表示無しでは，C，B が A に対してそれぞれ 4.9 倍，4.1 倍と高く評価され，A が最も低く評価された．音声相槌のみの提示の場合でも，うなずき反応モデルによる音声相槌の推定は応答として有効に機能していることが確認された．またキャラクタ表示有りでは，β と γ では差がなく α が最も低く評価された．

表 5.1　一対比較結果

	A	B	C	Total
A		4	5	9
B	20		10	30
C	19	14		33

キャラクタ表示なし

	α	β	γ	Total
α		10	8	18
β	14		13	27
γ	16	11		27

キャラクタ表示あり

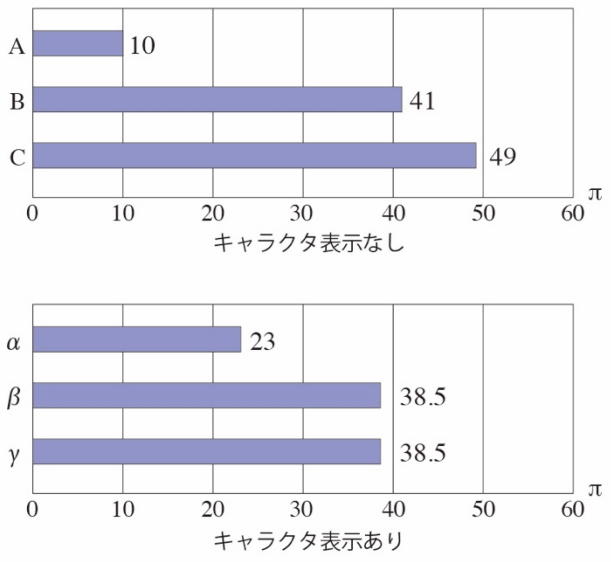

図 5.4　各モードの強さ π

全てのモードに対する 7 段階官能評価の結果について各項目の平均及び標準偏差を図 5.5 に示す．この結果に対して「キャラクタ表示の有無」と「音声相槌の有無とタイミング」の 2 つの要因に着目し，傾向の理解しやすさの観点から二要因分散分析を行った（表 5.2）．その結果すべての項目において，キャラクタ表示の有無と，音声相槌の有無とタイミングの 2 つの要因による主効果は共に有意水準 5 %で有意差が

認められ，本システムにおける自動生成による相槌提示の有効性が示された．2 つの要因の交互作用は認められなかった．

多重比較として Tukey の範囲検定を用いて分析を行った結果を図 5.5 に重ねて示す．うなずき反応モデルによる音声相槌を行う B，β モードと，音声相槌の出力タイミングを 300 ms 遅延させた C，γ モードとの比較では，全ての項目において B-C 間，β-γ 間にほとんど差が見られず，有意差も認められなかった．

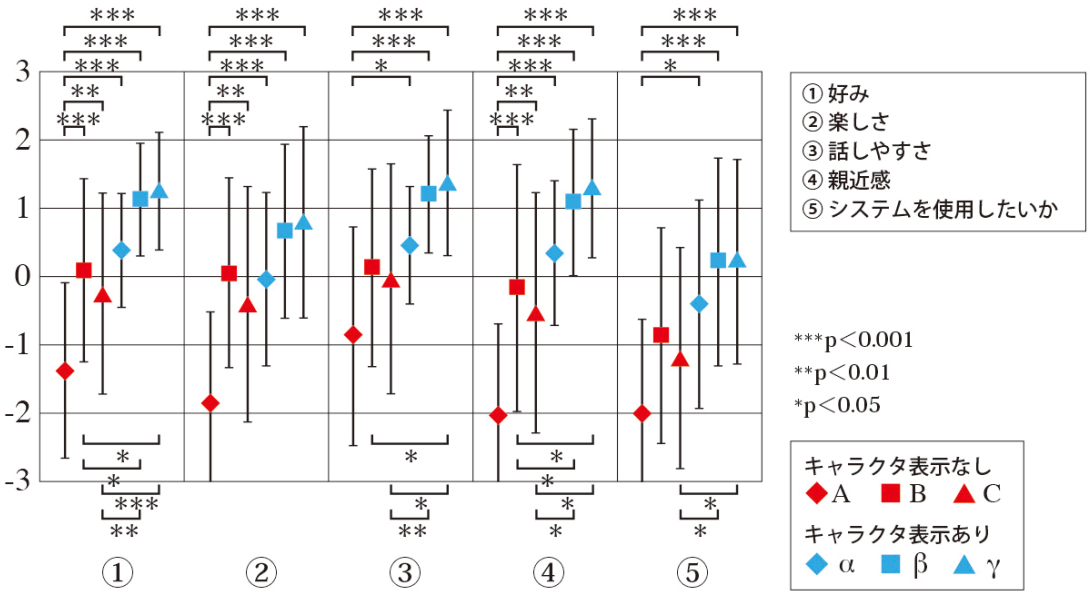

図 5.5　7 段階官能評価結果

表 5.2　二要因分散分析結果

① 好み

Source	degree of freedom	variation	variance	variance ratio	p-value
with character or not	1	73.67	73.67	50.655	0.000 ***
voice back-channel	2	35.68	17.84	12.266	0.000 ***
interaction	2	3.10	1.55	1.065	0.348
error		200.71	1.45		

② 楽しさ

Source	degree of freedom	variation	variance	variance ratio	p-value
with character or not	1	53.78	53.78	25.100	0.000 ***
voice back-channel	2	49.01	24.51	11.438	0.000 ***
interaction	2	8.76	4.38	2.045	0.133
error		295.67	2.14		

③ 話しやすさ

Source	degree of freedom	variation	variance	variance ratio	p-value
with character or not	1	58.78	58.78	30.522	0.000 ***
voice back-channel	2	24.50	12.25	6.361	0.002 **
interaction	2	0.72	0.36	0.188	0.829
error		265.75	1.93		

④ 親近感

Source	degree of freedom	variation	variance	variance ratio	p-value
with character or not	1	119.17	119.17	25.100	0.000 ***
voice back-channel	2	51.85	25.92	11.438	0.000 ***
interaction	2	7.60	3.80	1.745	0.178
error		300.38	2.18		

⑤ システムを使用したいか

Source	degree of freedom	variation	variance	variance ratio	p-value
with character or not	1	66.7	66.69	27.203	0.000 ***
voice back-channel	2	49.01	24.51	4.226	0.016 *
interaction	2	1.6	4.38	0.317	0.728
error		338.3	2.45		

実験時に得られた自由記述回答を表 5.3 に示す．肯定的意見は本システムによる反応によって，話しやすさや楽しいといったものが多かったが，「人相手には話しにくいようなことを話せる感じがあった」のように，対話エージェントとして新たな利用価値が認められる意見もあった．否定的意見としては，想定していないタイミングや不自然なタイミングで返された音声相槌によって発話を遮られ，話しにくさを感じるといったものが多く，改善の余地も多いことが分かる．

表 5.3　自由記述

肯定的な意見
・あいづちがすぐにあった方が自分の話についてきてくれているような感覚がしてよかった ・うなずきに加えて相槌を打ってくれる方が話しやすいと思った ・自分のしゃべったあとや，話の合間に良い間で，うなずきを入れてくれていたので話しやすく感じた ・人相手には話しにくいようなことを話せる感じがあった ・キャラクタがいる方が親近感がわきやすいのでいる方がいいと感じた ・相槌があった方が話しやすいと感じた．キャラクタはあった方が楽しいし話やすいと感じた ・キャラクタがいる方が安心する．言った後に音声とうなずきが入る方が好み ・キャラクタがいないと固いしゃべり方になった気がした
否定的な意見
・はやく返事されると少し話しにくかった ・相槌が多いと多少しゃべりづらいかなと感じた ・うなずきよりも声を発する方が話しやすいと思うこともあったが，タイミングがずれていると，逆に話しにくかった ・キャラ無し相槌無しだとただの独り言の気分でかえって話しやすかった ・キャラクタも音声も何もない状況だと自分の話を聞いてもらえているのかよくわからない，が反面，意識せずに話せるので緊張はしない ・短い区切り目のときに「うん」といわれたときちょっと話の途中で「ん？」となってしまうことがあった ・話している途中で「うん」とさえぎられると，少し話しにくいなと感じた

5.4.3 考察

実験結果より，キャラクタ表示有りのうなずき動作のみ（α モード）と，キャラクタ表示無しの音声相槌のみ（B および C モード）は同程度の評価であり，聞き手反応として評価されたと考えられる．うなずき動作と音声相槌を組み合わせた場合には，より豊かなコミュニケーション反応として知覚され，高く評価されている．うなずき動作に対する音声相槌のタイミングについては，うなずき動作のタイミングと同時に

発生するモードと 300 ms 遅れて発声するモードの間には評価の差は見られなかった．既述の通り，今回の音声相槌は長さ約 220 ms であり，うなずき動作が 500 ms である．つまり，音声相槌を遅らせず，同時に出力した場合はうなずきと音声相槌の開始が一致し，300 ms 開始を遅らせた場合はうなずきと音声相槌の終了がほぼ一致する状態となるため，違和感なくほぼ同程度の評価となったと考えられる．

5.5 音声相槌の頻度の検討

5.5.1 予備検討

対面コミュニケーションにおけるうなずき動作に対する音声相槌が伴う割合として，日本人同士で 70 %，アメリカ人同士で 42 %との報告がある［5.5］．そこで，音声相槌を伴わない（0 %），全てのうなずきに音声相槌を伴う場合（100 %）を加えた 4 つのモードで改めて予備検討を行った．20 歳～23 歳までの男女学生 11 名に対し，4 つのモードから 2 つのモードを抽出し，どちらが良かったかを聞く一対比較を 1 人につき 6 通り（${}_4C_2$）行った．一対比較の結果を表 5.4 に示す．Bradley-Terry モデルを想定し，各モードの強さ π を最尤推定した結果を図 5.6 に示す．うなずき動作に対して音声相槌が 70 %となるモードが 0%のモードに対して 13.2 倍，42 %のモードに対して 2.7 倍，100 %のモードに対して 2.2 倍と高く評価され，この結果より本システムではうなずき動作に対して音声相槌が 70 %となる閾値を設定した．

表 5.4　一対比較結果

	0%	42%	70%	100%	計
0%		2	1	1	4
42%	9		2	6	17
70%	10	9		7	26
100%	10	5	4		19

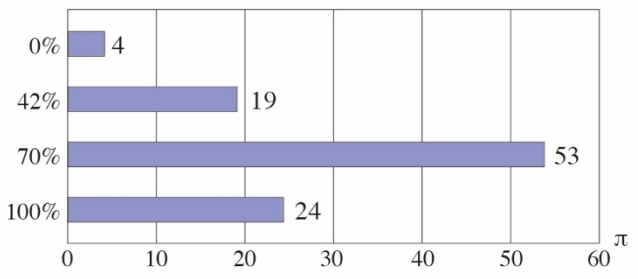

図 5.6　各モードの強さ π

5.5.2 実験方法

実験参加者にはディスプレイを見せながら対話を行わせた. 実験は自由な話題で語りかけをさせた．語りかけの話題が尽きないように 30 程度のテーマを提示し，それを参考に語りかけさせた．実験では次に示す I〜III の 3 つのモードを用意し，それぞれのモードについての評価を行った.

I ：従来のうなずきのみの InterActor（0 %）
II ：InterActor のうなずき全て（100 %）に「うん」と発声する相槌を付加
III：InterActor のうなずきの内，70 % に「うん」と発声する相槌を付加

実験参加者に，実施の流れ，注意事項，アンケート内容を説明した．その後ディスプレイの前に着席し，システムを実際に使用してみせて各モードの説明を行った．部屋に実験参加者だけを残し，I〜III のモードからランダムに 2 つのモードを抽出して順に使用させ，どちらが総合的に良いかをアンケートに答えさせた．比較するモードは 3 つある為，これを計 3 回（${}_3C_2$）繰り返した．次に，I〜III のモードを 1 つのモードにつき 1 分間使用させ，各セッション終了時に「楽しさ」「好み」「対話しやすさ」「親近感」「システムを使用したいか」の 5 項目について 7 段階（中立 0）で官能評価させた．また，自由記述欄にシステム使用中に気付いたことを全て記入させた．実験参加者は 19〜23 歳の男女学生 24 人で，実験参加者には 500 円の図書カードを謝礼として渡した.

5.5.3 実験結果

一対比較の評価結果を表 5.5 に示す．Bradley-Terry モデルを想定して強さ π を最尤推定した結果を図 5.7 に示す．その結果，III のモードが I のモードに対して 7.1 倍，II のモードに対して 2.3 倍と最も高く評価され，II のモードは I のモードに対して 3.0 倍高く評価された.

また 7 段階官能評価の結果について各項目の平均及び標準偏差を図 5.8 に示す．Friedman 検定を行った結果，全ての項目において有意水準 5 %で有意差が認められた

ため，Bonferroni による多重比較として Wilcoxon 符号順位検定を用いて各項目の有意差を求めた．評価において，I と II の間に「好み」「親近感」の項目において有意水準 1 %で有意差が認められ，「楽しさ」「対話しやすさ」「システムを使用したいか」の項目において有意水準 5 %で有意差が認められた．I と III の間では「システムを使用したいか」の項目において有意水準 5 %で有意差が認められ，「好み」「楽しさ」「対話しやすさ」「親近感」の項目において有意水準 1 %で有意差が認められた．

実験時に得られた自由記述回答を表 5.6 に示す．

表 5.5　一対比較結果

	I	II	III	計
I		7	2	9
II	17		8	25
III	22	16		38

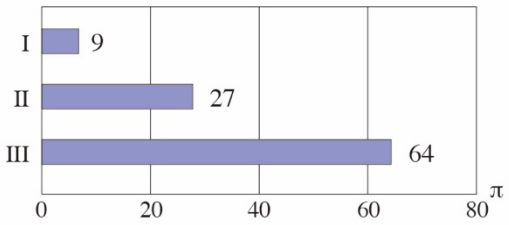

図 5.7　各モードの強さ π

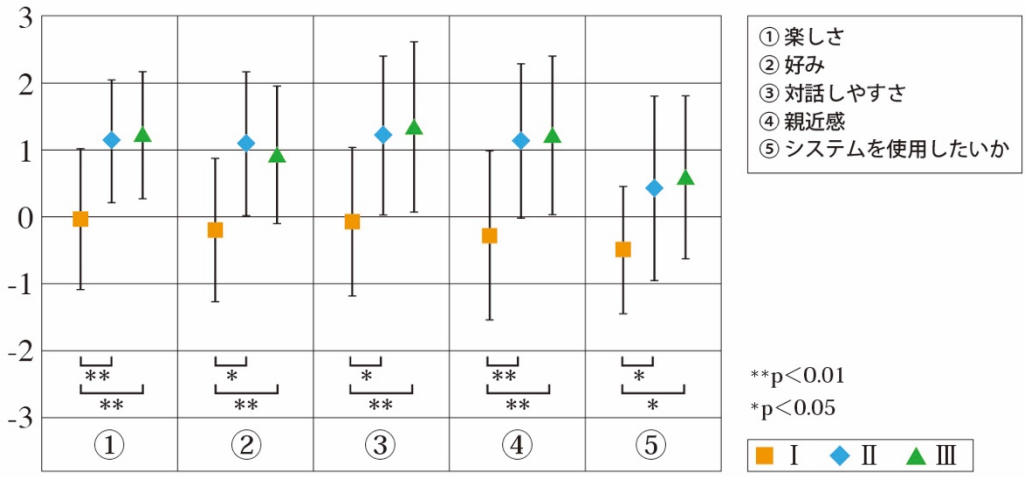

図 5.8　7 段階官能評価結果

表 5.6　自由記述

肯定的な意見
・相手がいないので話すのに抵抗があったが，キャラクタが反応してくれるので楽しめた ・相槌を良くしてくれているときは話しかけやすかった ・会話に合わせて相槌をしてもらえると，非常に話しやすいと感じた ・相槌があると話していて楽しくなった ・Ⅱ，Ⅲが話しやすく，親近感がわく気がした ・頷いてくれるだけよりも「うん」と言ってくれている方がより話を聞いてくれるという感じがしてよかった ・声があることで，さらに話そうという気になった ・うなずき動作と音声を組み合わせると話しやすく，会話を楽しめた ・Ⅱは空気を読まずに「うんうん」言ってくれるのでよかったように感じる ・声が出るのが最初はあまり好きではなかったが，なれると声が出た方がいいなと思った ・毎回返事が返されるよりは，必要に応じて返事をしてくれた方が自分としては話しやすかった
否定的な意見
・声が返ってこないと少し不安に感じてしまった ・Ⅰは一人で話している感じがして寂しかった ・「うん」だけでなく別の反応があるといいと思った ・ⅡとⅢの違いがあまりわからなかった ・一方的にこちらが話しているだけなので，Ⅲだと無音の時が長く感じて早く「うん」と言ってほしい気持ちになった

5.5.4 考察

一対比較の結果，対話中のうなずき動作に対し 70 %の生起確率で音声相槌を付加する III モードが最も高く，次にすべてのうなずき動作に対し音声相槌を付加する II モードが高く評価されたことから，音声相槌の生起確率は人同士の応答特性を考慮する必要があることが考えられる．また，7 段階官能評価においても I と II の間，I と III の間のすべての項目においてそれぞれ音声相槌があるモードが高く評価され，有意差が認められたことから，音声相槌をうなずき動作へ付加することが高く評価されたと考えられる．しかしながら 7 段階官能評価では II と III の間においてどの項目にも有意差が認められなかった．その理由として，参考にした音声相槌の特性は人同士の対面コミュニケーションでお互いに話し合っている場合での特性であり，今回の実験では InterActor の CG キャラクタは聞き手役のみを担っていたためと考えられる．

自由記述では「声があることでさらに話そうという気持ちになった」，「うなずき動作と音声を組み合わせると話しやすく，会話を楽しめた」等の意見も得られたこと

から, 音声相槌は使用者の発話意欲を引き出す効果があると考えられる. 一方で「「うん」だけではなく別の反応があるといいと思った」という意見もあったことから，文脈によって音声相槌の発生確率や音声応答の内容を変更することで評価が異なる可能性があり，また対話相手としてのエージェント機能拡充が期待される.

5.6 おわりに

本章では, うなずき動作に音声相槌を伴う音声駆動型身体的引き込みキャラクタシステムを開発し，キャラクタ表示無しにおいて音声相槌そのものの評価を，またキャラクタ表示と組み合わせた音声相槌の効果を, それぞれ語りかけ実験によって評価した．その結果，音声相槌の提示は高く評価され，うなずき動作に音声相槌を伴う提示がより高く評価された. また日本語対話での応答特性を考慮した頻度で音声相槌を自動生成するキャラクタシステムを開発し, 評価実験によってシステムの有効性を示した.

参考文献

［5.1］ 川嶋宏彰，スコギンズ・リーバイ，松山隆司，漫才の動的構造の分析―間の合った発話タイミング制御を目指して―，ヒューマンインタフェース学会, Vol.9, No.3 (2007), pp.379–390.

［5.2］ A. Pomerantz, Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes, J. M. Atkinson and J. Heritage, Cambridge University Press (1984), pp.57–101.

［5.3］ 須藤潤, 日本語感動詞「うん」の意味・機能の分類から音声的特徴の分析へ, 音声研究, Vol.11, No.3 (2007), pp.94–106.

［5.4］ 石井カルロス寿憲, 音声対話中に出現するパラ言語情報と音響関連量：声質の役割に焦点を当てて（＜小特集＞音声は何を伝えているか），日本音響学会誌, vol.71, No.9 (2015), pp.476–483.

［5.5］ 泉子・K・メイナード, 会話分析, くろしお出版 (1993), pp.54–59.

［5.6］ 山本倫也, 渡辺富夫, ロボットとのあいさつインタラクションにおける動作に対する発声遅延の効果, ヒューマンインタフェース学会論文誌, Vol.6, No.3 (2004), pp.87–94.

第 6 章

感情極性に基づく音声相槌システム

6.1 はじめに

前章では, 音声駆動型身体的引き込みキャラクタのうなずき動作に音声相槌を付加したシステムを開発し, うなずき動作と音声相槌を組み合わせた提示手法が好ましいという結果が得られた.

本章では, 前章で開発したうなずき動作に音声相槌を伴う音声駆動型身体的引き込みキャラクタシステムにおいても, 負の感情を助長する可能性を回避する手法を検討する．前述の通り，音声相槌を少し遅らせることで意味的な解釈を与え，使用者の発話感情に対応したシステムとして活用できる可能性がある．そこで，まず前章で開発したシステムを用いて, うなずき動作に対する音声相槌の提示タイミングを検討するための評価実験及び Negative な発話が行われた際の音声相槌の提示タイミングを検討するための評価実験を行った．その結果，InterActor が推定したうなずき動作の開始から 600 ms 程度までの音声遅延が許されることが示された．また，Negative な発話内容に対する音声相槌のタイミングは 900 ms 程度の遅延まで許容されることが示唆された．次に，これらの知見をシステム統合することで，音声認識により得られた発話内単語の感情極性に基づいた, 音声相槌を伴う身体的引き込みキャラクタシステムの開発を行った.

6.2 うなずき動作に対する音声相槌タイミングの評価実験

6.2.1 実験方法

うなずき動作に伴う音声相槌の提示タイミングの効果を検討するための評価実験を行った．うなずき動作の出力タイミングは 2.4 節で述べた通り，実際の発話終了から 133 ms のハングオーバー処理を施された音声データを用いて，iRT により予測する．実験は，5.4.2 項の実験でキャラクタ表示ありの状態で遅延なし（うなずき反応モデルによる音声相槌）と 300 ms 遅延させた音声相槌の差がなく，かつ先行研究[6.1]で動作開始から 300 ms の遅延が丁寧で評価が高いことから 300 ms 遅延させた音声相槌を基準となる A モードとして，次に示す提示タイミングの異なる 4 つのモードを用意した．

A：InterActor が推定したうなずき動作の開始から 300 ms のタイミングで音声相槌を行う

B：InterActor が推定したうなずき動作の開始から 600 ms のタイミングで音声相槌を行う

C：InterActor が推定したうなずき動作の開始から 900 ms のタイミングで音声相槌を行う

D：InterActor が推定したうなずき動作の開始から 1200 ms のタイミングで音声相槌を行う

相槌の発声の有無による影響を無くすため，Threshold と Threshold A を同じ閾値としてキャラクタは全てのうなずき動作に対して相槌を行う．まず，4 つのモードから 2 つのモードを抽出した一対比較を 1 人につき 6（${}_4C_2$）通り行った．次に各モードを使用させ，「好み」「楽しさ」「話しやすさ」「親近感」「丁寧さ」「システムを使用したいか」の 6 項目について 7 段階官能評価（中立 0）を行った．また，自由記述欄に気づいたことを記入させた．1 つのモードの使用時間は一対比較では 1 分間，7 段階官能評価では 1 分 30 秒間とし，実験参加者は自由なテーマによる語りかけを行った．モードの提示順序は順序効果を考慮して，カウンターバランスをとった．実験

参加者は 19～24 歳の男女各 12 名の計 24 名で，実験参加者には 500 円の図書カードを謝礼として渡した．

6.2.2 実験結果

一対比較の結果を表 6.1 に示す．一対比較による評価を一義的に定めるために，Bradley-Terry モデルを想定し，モードの強さ π を最尤推定した結果を図 6.1 に示す．その結果，A，B，C，D の順で高く評価され，A と B は，C と D に対し 2 倍以上の強さであった．

表 6.1　一対比較結果

	A	B	C	D	計
A		12	20	17	49
B	12		16	18	46
C	4	8		14	26
D	7	6	10		23

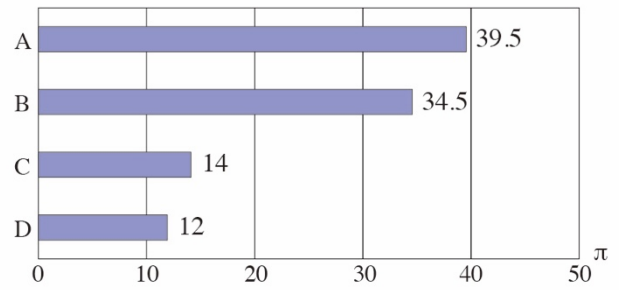

図 6.1　各モードの強さ π

7 段階官能評価の結果について各項目の平均及び標準偏差を図 6.2 に示す．各項目間の比較を行うために Wilcoxon の符号順位検定を行った．その結果，「好み」の項目で A と D の間に有意水準 5 %，B と D の間に有意水準 1 %，「楽しさ」の項目で A と D の間に有意水準 1 %，B と D の間に有意水準 5 %，「話しやすさ」の項目で A と D の間に有意水準 1 %，B と D の間に有意水準 0.1 %，「親近感」の項目で A と D の間に有意水準 1 %，B と D の間に有意水準 5 %，「システムを使用したいか」の項目で A と D，B と D の間に有意水準 0.1 %で有意差が認められた．また，「話しやすさ」の項目において C と D の間に，「親近感」の項目において A と C の間に，「システムを使用したいか」の項目において B と C の間に有意水準 5 %で有意差が認められた．「丁寧さ」の項目においてはいずれも有意差が認められなかった．

実験時に得られた自由記述回答を表 6.2 に示す．「ちゃんと話が区切られているところで返事が返ってきたので話しやすかった」「話を聞いてもらえていると感じた」などの肯定的意見と，「話している途中に相槌が入ったりして話しにくいモードがあった」「頷きと声が遅れると不快に感じる」「相槌が遅れて出てくるモードは変な丁寧さがあった」などの否定的意見が見られた．B モードでは肯定的なコメントが，C モードでは否定的なコメントが多く見られた．

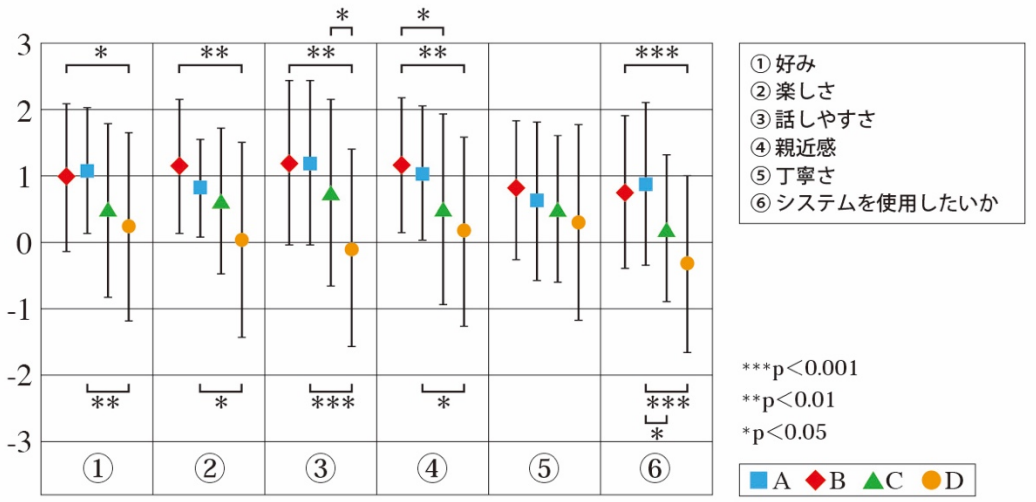

図 6.2　7 段階官能評価結果

表 6.2　自由記述

肯定的な意見
・良いタイミングで「うん」が返ってくると次も話しやすい ・慣れてくると独り言も楽しかった ・話を聞いてもらえていると感じた．もっと精錬されてバリエーションが豊かになれば心を癒せるものになると思う ・ちゃんと話が区切られているところで返事が返ってきたので話しやすかった．返事だけでなく頷く仕草も良かった ・個人的には返事が返ってくるのが早い方が好みですが，遅い方が丁寧だと思った
否定的な意見
・頷きと「ウン」の間隔，モードごとに違いが小さいので，一対比較がちょっと難しかった ・話している途中に相槌が入ったりして話しにくいモードがあった ・頷きと声が遅れると不快に感じる ・人と話していると思うと違和感がある ・相槌が遅れて出てくるモードは変な丁寧さがあった

6.3 シナリオに基づく Negative な語りかけによる評価実験

6.3.1 実験方法

前節の評価実験により，InterActor が推定したうなずき動作の開始から 600 ms 程度までの音声遅延が許されることを示した．しかし，使用者の発話内容によっては，適切な音声相槌のタイミングが異なる可能性がある．そこで，シナリオに基づいた語りかけによる実験を行い，Negative な発話が行われた際の音声相槌の好ましい提示タイミングを検討するための評価実験を行った．実験は，次に示す音声相槌の提示タイミングの異なる 3 つのモードを用いた．

α：InterActor が推定したうなずき動作の開始から 300 ms のタイミングで音声相槌を行う

β：InterActor が推定したうなずき動作の開始から 600 ms のタイミングで音声相槌を行う

γ：InterActor が推定したうなずき動作の開始から 900 ms のタイミングで音声相槌を行う

相槌の発声確率は日本語話者の相槌の特性に基づいて 70 %とした．音声ピッチについては, 基本周波数が高い場合は驚きや喜びなどの快感情と認知されることが報告されていることから［6.2］，Negative な発話内容に対して不適切な応答になることを避けるために，前章の実験で使用したものよりもピッチを下げた 338 Hz とした．まず，　3 つのモードから 2 つのモードを抽出した一対比較を 1 人につき 3（$_3C_2$）通り行った. 次に各モードを使用させ, 前章の実験において有意差の認められなかった「丁寧さ」を除外し，Negative な発話内容における評価のために「安心感」「和み」を加えた 7 項目（好み，楽しさ，話しやすさ，親近感，安心感，和み，システムを使用したいか）について 7 段階官能評価（中立 0）を行った．また，自由記述欄に気づいたことを記入させた．1 つのモードの使用時間は一対比較では 60 秒間，7 段階官能評価では 90 秒間とした．モードの提示順序は順序効果を考慮して，カウンターバランスをとった．実験参加者は 19～24 歳の男女各 12 名の計 24 名で，実験参加者には 500

円の図書カードを謝礼として渡した. 使用したシナリオは 3 章で使用したものと同様である（図 3.1）.

6.3.2 実験結果

一対比較の結果を表 6.3 に示す．一対比較による評価を一義的に定めるために，Bradley-Terry モデルを想定し，モードの強さ π を最尤推定した結果を図 6.3 に示す．その結果，α，γ，β の順で高く評価され，各モード間にほとんど差は認められなかった．

表 6.3　一対比較結果

	α	β	γ	計
α		12	13	25
β	12		11	23
γ	11	13		24

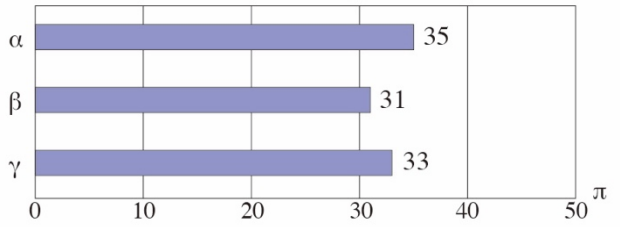

図 6.3　各モードの強さ π

7 段階官能評価の結果について各項目の平均及び標準偏差を図 6.4 に示す．各項目間の比較を行うために Wilcoxon の符号順位検定を行った．α と γ の間に，「好み」の項目において有意水準 1 %で，「楽しさ」「話しやすさ」「システムを使用したいか」の項目において有意水準 5 %で有意差が認められた．「親近感」「安心感」「和み」の項目においてはいずれも有意差が認められなかった.

実験時に得られた自由記述回答を表 6.4 に示す．肯定的意見は反応があることによって話しやすさを感じたというものが多く，否定的意見は「うなずきの音声が遅いので違和感があった」「相槌が話の途中で入ってきて「おや」と感じた」などのタイミングに関する意見や音声の衝突によって発話が阻害される点についての意見が見られた.

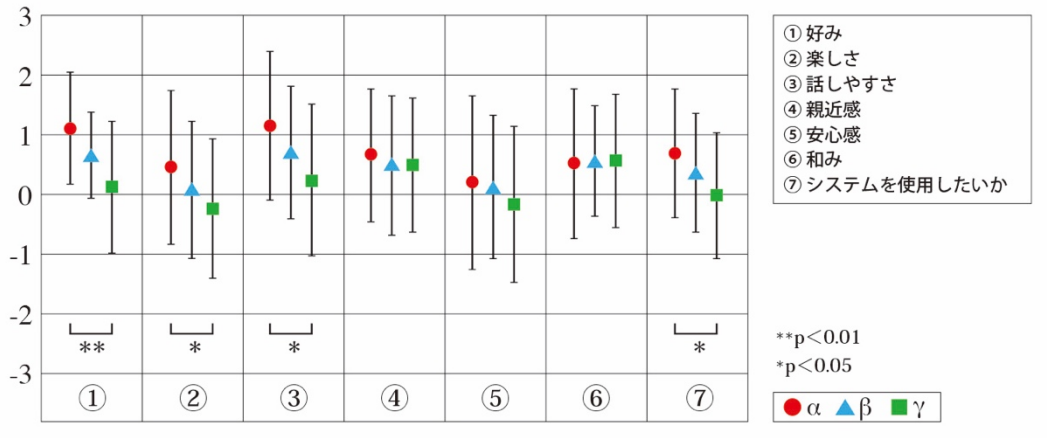

図 6.4　7 段階官能評価結果

表 6.4　自由記述

肯定的な意見
・テンポよく話を進められた ・ちゃんと話の内容を理解して同情してくれているような感じがした ・考えながらゆっくりと話をすることができた ・少し冷たさを感じたが，しっかり話を受け止めてくれているイメージを受けた ・キャラクタがいる方が親近感がわきやすいのでいる方がいいと感じた ・自然な相槌に感じた
否定的な意見
・相槌の速さが少し早かったように感じ，あまり親近感や話しやすさは感じなかった ・ちゃんと話を聞いているのかという気分になった ・何か別の作業や考え事をしながらとりあえずは相槌を返してくれているような印象を受けた ・うなずきの音声が遅いので違和感があった ・相槌が話の途中で入ってきて「おや」と感じた

6.3.3 考察

6.2 節の結果より，InterActor が推定したうなずき動作の開始から 300 ms と 600 ms のタイミングで音声相槌を行うモードに差が見られず，600 ms と 900 ms のタイミングとの間には大きな差が確認されたことから，うなずき動作に対して音声相槌を 600 ms 程度まで遅延させて提示することが許容されることを示した．一方，6.3 節の一対比較結果より，6.2 節において評価の低かった 900 ms のタイミングの音声相槌が，300 ms 及び 600 ms のタイミングの音声相槌と同程度評価されていることから，Negative な発話内容に対する音声相槌のタイミングは 900 ms 程度の遅延まで許容されることが示された．7 段階官能評価の結果からは，「好み」「楽しさ」「話しやすさ」「システムを使用したいか」の項目で 900 ms のタイミングの音声相槌が低く評価されたが，「親近感」「安心感」「和み」の項目では差が見られなかったことからも，Negative な発話内容に対する聞き手役としての可能性が示唆された．これらの結果は，使用者の負の感情での発話に対して，あえて音声対話エージェントの音声反応を遅らせることで否定的な反応として利用するなどの応用が可能であると考えられる．

6.4 感情極性に基づく音声相槌システムの構築

これまでの知見を統合し，音声認識により得られた発話内単語の感情極性に基づいて，Negative な発話内容の場合に応答遅延を伴う音声相槌を行う身体的引き込みキャラクタシステムを開発した．キャラクタは話者の語りかけに対して自動生成される身体的引き込み動作に加えて，音声相槌を行う．Negative な発話内容の場合は InterActor が推定したうなずき動作の開始から 900 ms のタイミングで，それ以外の場合は 300 ms のタイミングで音声相槌を行う．使用者の負の感情での発話に対して，あえて音声相槌を遅らせることで，キャラクタが使用者に不快感を与えることなく否定的な反応を示すことができる可能性がある．システムのコンセプトを図 6.5 に示す．

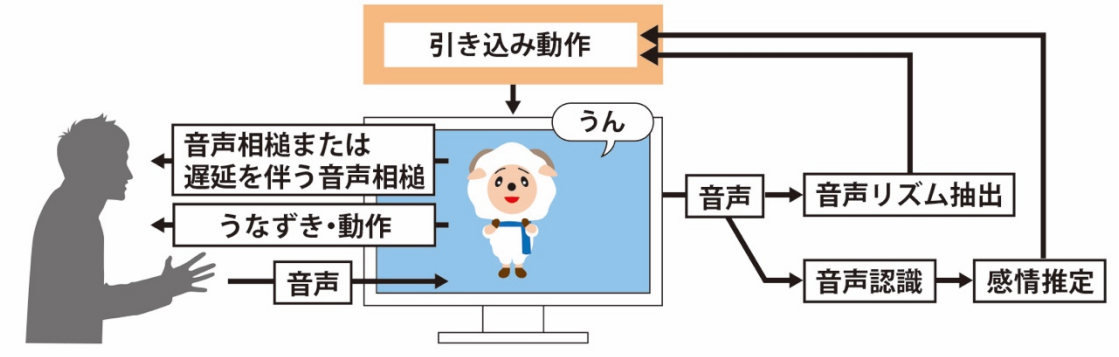

図 6.5　コンセプト

なお，開発したシステムは音声相槌のみに着目したシステムである．2 章では発話内単語の感情極性に基づいて使用者の発話感情を推定し，それに対応した反応動作を提示するシステムを開発したが，動作と音声相槌を組み合わせた場合，提示される動作の種類によっては「うん」という音声が整合しないと感じられる可能性があるため，発話内単語の感情極性に基づいて音声相槌タイミングを変化させるシステムとなっている．

開発環境は Windows 7 32 bit PC（HP Core i5 CPU, 2.67 GHz, 4 GB メモリ），DirectX 及び Visual Studio で，開発したシステムを使用している様子を図 6.6 に示す．

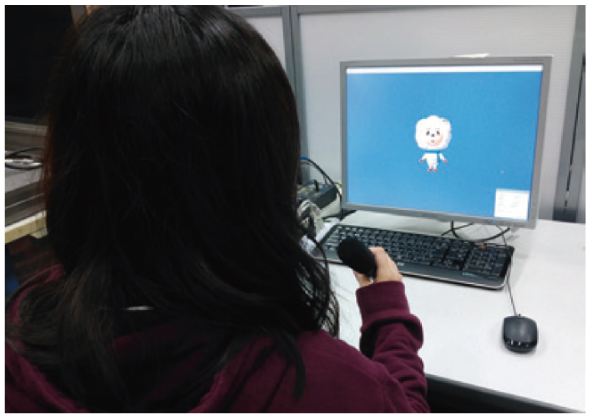

図 6.6　システム使用の様子

6.5 おわりに

本章では，うなずき動作に音声相槌を伴う音声駆動型身体的引き込みキャラクタシステムのうなずき動作に対する音声相槌の好ましい提示タイミングを検討するための評価実験及び Negative な発話が行われた際の音声相槌の提示タイミングを検討するための評価実験を行った．その結果，InterActor が推定したうなずき動作の開始から 600 ms 程度までの音声遅延が許されることが示された．また，Negative な発話に対する音声相槌のタイミングは 900 ms 程度の遅延まで許容されることが示唆された．

次に，うなずき動作に音声相槌を伴う音声駆動型身体的引き込みキャラクタシステムに，発話内単語の感情極性から話者の状態を推定する状態推定モデルを導入し，音声認識により得られた単語感情極性に基づいて，Negative の場合に 900 ms の応答遅延を伴う音声相槌を行う身体的引き込みキャラクタシステムの開発を行った．

今後の課題として，システムを用いた評価実験を行い，システムの有効性を確認していく必要がある．また，本システムは音声情報を用いて全ての応答動作を自動生成するため，マイクと PC のみで構成できる応用可能性が高いシステム構成となっている．しかし韻律情報と言語情報のそれぞれから得られる情報については，人間は対話の印象を主に韻律情報で得ており，言語情報の役割は小さい［6.3］との指摘もあり，今後さらに詳細に検討する必要がある．

参考文献

［6.1］ 山本倫也，渡辺富夫，ロボットとのあいさつインタラクションにおける動作に対する発声遅延の効果，ヒューマンインタフェース学会論文誌，Vol.6，No.3 (2004)，pp.87–94.

［6.2］ 重野純，感情を表現した音声の認知と音響的性質，心理学研究，Vol.74，No.6 (2004)，pp.540–546.

［6.3］ 西村良太，北岡教英，中川聖一，音声対話における韻律変化をもたらす要因分析，音声研究，Vol.13，No.3 (2009)，pp.66–84.

第 7 章

結論

7.1 本研究のまとめ

本研究では, 音声入力からうなずきなどのコミュニケーション動作を自動生成する音声駆動型身体的引き込みキャラクタ InterActor に対して, 身体的リズム同調を損なうことなく, 使用者の発話感情に応じた反応動作または音声相槌を提示するシステムを開発し，その評価を行い，有効性を示した．以下，本論文における各章の成果をまとめる．

2 章では，従来の音声駆動型身体的引き込みキャラクタ InterActor に音声認識を導入し，話者の発話内単語の感情極性から使用者の状態を推定し，結果に基づいて反応動作を変化させる音声駆動型身体的引き込みキャラクタシステムを開発した．

3 章では，2 章で開発したシステムの評価実験を行った．まず，Negative または Positive と判定されたシナリオに基づくキャラクタへの語りかけによる自動応答エージェントシステムとしての評価実験を行った．その結果，一対比較では提案手法を用いたモードが高く評価された．7 段階官能評価では，Negative と判定されたシナリオに基づく語りかけ実験において, 従来の InterActor の聞き手動作を行うモードに対し, 提案手法を用いたモードで「楽しさ」「好み」「安心感」「和み」の項目で有意差が確認された．Positive と判定されたシナリオに基づく語りかけ実験においては，全て

の項目（楽しさ，好み，対話しやすさ（話しやすさ），安心感，和み，システムを使用したいか）で有意差が確認された．次に二者対話によるコミュニケーションインタフェースシステムとしての評価実験を行った．その結果，一対比較では提案手法を用いたモードが高く評価され，7 段階官能評価においても全ての項目で有意差が確認され，システムの有効性が示された．

4 章では，使用者の多様な感情に対応することを目的に，2 章で開発したシステムに発話時間率に基づく活性度を加えて話者の状態を推定する状態推定モデルを定義し，結果に基づき反応動作を行う身体的引き込みキャラクタシステムを開発した．発話活性度を考慮した状態推定の有効性を検討するために 2 章で開発したシステムとの比較実験を行ったが，評価実験における活性度の状態変化が小さいことでモードの判別がつきにくかった可能性もあり，十分な効果は認められなかった．

5 章では，まず従来の音声駆動型身体的引き込みキャラクタ InterActor に音声相槌を付加したキャラクタシステムを開発し，自動生成する音声相槌についてその出力タイミングとキャラクタ表示による効果に着目した評価実験を行った．その結果，音声相槌の提示は高く評価され，音声相槌のみの提示の場合でも，うなずき反応モデルによる音声相槌の推定は応答として有効に機能していることが確認された．また，キャラクタ表示有りでは，うなずき動作に音声相槌を伴うモードがうなずき動作のみのモードよりも高く評価された．次に，日本語対話での応答特性を考慮した頻度で音声相槌を自動生成するキャラクタシステムを開発して評価実験を行った．その結果，一対比較では提案手法を用いたモードが最も高く評価された．また，7 段階官能評価では全ての項目（楽しさ，好み，対話しやすさ，親近感，システムを使用したいか）で有意差が確認され，システムの有効性が示された．

6 章では，5 章で開発したうなずき動作に音声相槌を伴う音声駆動型身体的引き込みキャラクタシステムにおいても，否定的な発話感情を助長する可能性を回避する手法を提案した．まず，うなずき動作に対する音声相槌の好ましい提示タイミングを検討するための評価実験及び Negative な発話が行われた際の音声相槌の提示タイミングを検討するための評価実験を行った．その結果，InterActor が推定したうなずき動作の開始から 600 ms 程度までの音声遅延が許されることが示された．また，Negative

な発話内容に対する音声相槌のタイミングは 900 ms 程度の遅延まで許容されることが示唆された．次に，うなずき動作に音声相槌を伴う音声駆動型身体的引き込みキャラクタシステムに, 発話内単語の感情極性から話者の状態を推定する状態推定モデルを導入し，音声認識により得られた発話内単語の感情極性に基づいて，Negative の場合に 900 ms の応答遅延を伴う音声相槌を行う身体的引き込みキャラクタシステムを構築した.

7.2 今後の展望

本研究で開発したシステムは, エージェントが使用者の否定的な発話感情を肯定し, 助長しないという観点から, 身体的引き込みと使用者の感情推定に基づく応答により, 人と対話エージェントのインタラクションを円滑にし, よりよい関係にしていくものである．今後様々な情報システムと融合することで付加価値の高いシステム，技術として発展・進化し，産業的発展も期待される．しかし，使用者が文脈によって正負の意味合いが変わるような単語を用いた発話を行った場合, 本研究で採用した単語による状態推定というアプローチで十分とは言えない．今後，単語以外の情報を含めて使用者の状態を推定することが大きな課題である．さらに，適切な応答としての動作表現や音声相槌のあり方を詳細に検討していくことも必要である．

謝辞

本研究を行うにあたり，終始ご指導，ご鞭撻を賜りました岡山県立大学情報工学部教授，渡辺富夫博士に深く感謝いたします．また，本研究に有益なご助言，ご指導いただきました岡山県立大学情報工学部准教授，石井裕博士に深く感謝いたします．

さらに, 本研究を遂行する上でご協力いただいた岡山県立大学情報工学部ヒューマンインタフェース研究室の皆さんに深く感謝いたします．

各種実験に実験参加者としてご協力いただいた岡山県立大学の多くの学生の皆様に感謝の意を表します．

尚，本研究の一部は，科学研究費 JP16K00278，19K12067，ウエスコ学術振興財団の助成によるものであり，関係各位に感謝の意を表します．

本論文に関する研究業績

原著論文

［1］ 西田麻希子，太田靖宏，渡辺富夫，石井裕：発話内単語の感情極性に基づき反応動作を行う身体的引き込みキャラクタシステム，日本機械学会論文集，Vol.83，No.846 (2017)，pp.1–14，DOI: 10.1299/transjsme.16-00148.

［2］ 西田麻希子，渡辺富夫，石井裕：音声相槌を行う音声駆動型身体的引き込みキャラクタシステム，日本機械学会論文集，Vol.85，No.880 (2019)，pp.1–12，DOI: 10.1299/transjsme.19-00159.

国際会議議事録

［1］ Yutaka Ishii，Makiko Nishida，Tomio Watanabe：Development of a Speech-driven Embodied Entrainment Character System with a Back-channel Feedback，Advances in Affective and Pleasurable Design，AHFE 2018，Advances in Intelligent Systems and Computing，Vol.774 (2018)，pp.132–139.

口頭発表

［1］ 西田麻希子，渡辺富夫，石井裕：発話活性度および感情極性に基づき反応動作を行う身体的引き込みキャラクタシステムの開発，日本機械学会 2018 年度年次大会講演論文集(2018)，G1200101，pp.1–5.

［2］ 西田麻希子，渡辺富夫，石井裕：身体的引き込みキャラクタシステムにおける

音声相槌タイミングの評価, 日本機械学会 2019 年度年次大会講演論文集(2019), S12102，pp.1–5.

［3］ 西田麻希子，渡辺富夫，石井裕 : 発話内単語の感情極性に基づく音声相槌を伴う身体引き込みキャラクタシステムの開発，第 169 回ヒューマンインタフェース学会研究会研究報告集(2019)，pp.21–24.