

論文内容の要旨

論文題目 **Designing Multimodal Conversational Interfaces
Based on Empirical Studies of Human Communicative Behaviors**

(和訳) 人間のコミュニケーション行動に関する実証的研究に基づく
マルチモーダル会話インタフェースのデザイン

氏名 中野 有紀子

音声言語、画像、アニメーションに等、複数のコミュニケーションモダリティを有するマルチモーダル会話インタフェース (MCI) はその表現の豊富さから、よりわかりやすいメディアであると思われがちであるが、人間の認知やコミュニケーションの様式に適合した表現を適切に選択できなければ、その複雑さゆえにかえってわかりにくいものになってしまう。

本論文では、人間のコミュニケーション行動を解明するためのコミュニケーション科学、心理学、言語学からの知見、および方法論と、会話インタフェースを構築するための人工知能やメディア技術等の工学的なアプローチとを統合することにより、人間の言語・非言語コミュニケーション行動の分析に基づいた MCI のデザインを行い、より自然な人対コンピュータのインタラクションを実現することを目的とする。第 1 章では、本論文の研究方法について議論し、(1)人間のコミュニケーション行動を分析し、(2)それをモデル化する、(3)そのモデルをシステムに実装することにより MCI を構築し、それを評価する、という 3 つのステップからなる研究方法を提案する。このように実証研究とシステムデザインとを一貫した枠組みで遂行することにより、人間と同様のコミュニケーション様式を MCI の機能として実現することが可能となる。本論文の各章では、この研究方法を用いて MCI の各構成素をデザインし、構築する。簡略化した MCI のシステム構成とこれに対応付けた本研究の概要を図 1 に示す。

第 2 章では、本研究に関連する言語学、計算言語学、コミュニケーション科学、人工知能、会話エージェント等の研究を概観する。

第 3 章では、会話インタフェースにおけるシステム発話の決定 (Content Planner) に関する問題として、システムがユーザから情報を引き出すための質問の生成と、ユーザへの教示の生成を取り上げる。

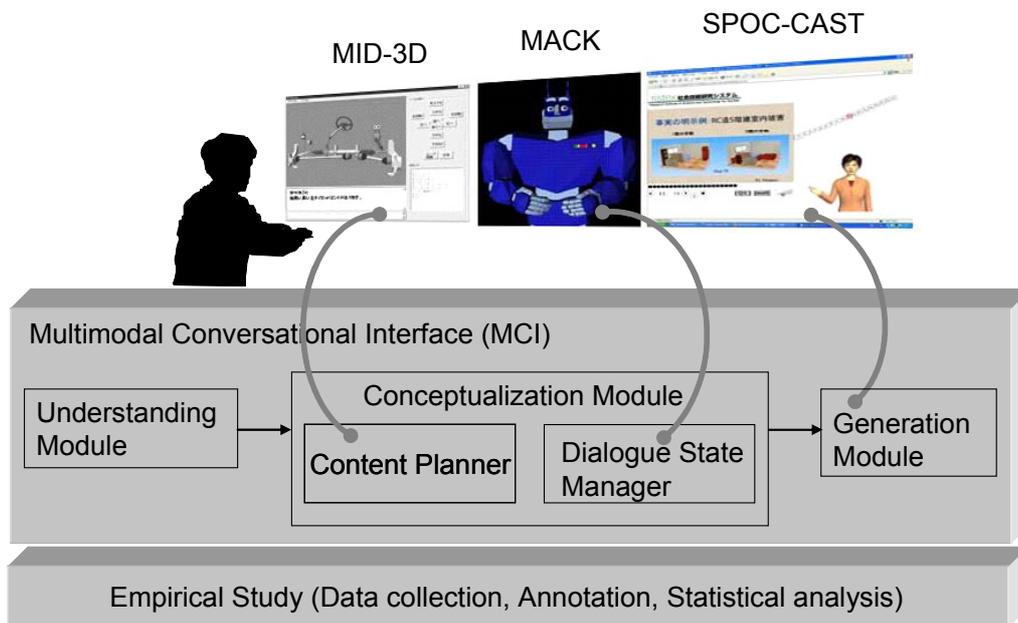


図 1 : MCI システム構成と本研究の概要

質問の生成についての研究では、まず質問内容を特徴付ける 5 つのパラメータを設定し、これらのパラメータ値の集合と質問の仕方との関係を調べた。この分析結果に基づき、質問内容のパラメータ値から質問の仕方を自動的に決定する機構を提案し、さらに評価実験により本手法の有効性を確認した。

対話的な教示の生成についての研究では、電話機操作の教示対話を分析し、初めての説明と、復習としての説明、学習者が理解できなかった場合の再説明の 3 つの対話状況で、付加的な説明の出現頻度と 1 回のターンで与える情報量が異なっていることを見出した。さらに、これらの点に着目して、教示の発話内容決定とターン方略の決定メカニズムをプログラミングを用いて実装した。

第 4 章では、第 3 章での言語のみの会話インタフェースを拡張し、3次元仮想環境を利用した教示対話システムにおいて、ユーザの視野、視点に応じた教示対話を生成する機構を提案する。

第 1 に、本システムでは、談話中の参照物が現在のユーザ視点から知覚可能であるか否かを制約として発話内容プランの展開を行うことにより、ユーザの視点に応じた発話内容の決定を行う。第 2 に、ユーザの視点对話履歴の一部として談話構造モデルを用いて管理することにより、話題の転換 (push)、回帰 (pop) が起こった場合にも、ユーザの視点を適切にコントロールすることが可能である。本システムは、自動車の構造を学習するための仮想環境であり、ユーザは音声と GUI によりシステムとのインタラクションを行う。

本章で提案した手法により、マルチモーダル教示対話システムにおいて、ユーザの視点

という知覚的な制約を考慮した対話の内容決定を行い、視覚情報と言語情報とを対話生成メカニズムの中に齟齬なく統合することが可能となった。

第 5 章では、MCI の対話制御部 (Dialogue State Manager) について述べる。また、ユーザとのより自然なインタラクションを実現するために、会話エージェントを導入したシステムを実装する。

会話の最も基本的な側面の 1 つとして、相手の言ったこと、意味したことを共通の理解とする grounding という過程がある。特に対面場面では、言語情報のみならず非言語情報もこの grounding 過程に寄与していると考えられる。そこで、本章では、うなづきや視線といった非言語情報の使用について grounding の視点から詳細に分析し、grounding の対象となる言語情報と grounding 過程で用いられる非言語情報との関係を探る。

まず、地図を共有しながら道案内をする対面場面の会話において、視線やうなづきがどのように使用されているのかを分析した結果、grounding の対象となる発話行為のタイプによって、聞き手からの理解の証拠 (evidence of understanding) となる非言語情報の使用パターンが異なっていることが明らかになった。さらに、聞き手の継続的な注視行動は理解の負の証拠 (negative evidence of understanding)、つまり、聞き手が理解できなかったことを示す非言語情報であることがわかった。

以上の分析結果に基づき、grounding ための非言語情報を生成、認識できる会話エージェントを実装した。予備的な評価実験として、ユーザが本システムとインタラクションする場合と、grounding 機能を持たないシステムとインタラクションする場合とを比較した。その結果、本システムとのインタラクションにおけるユーザの非言語行動が人対人のインタラクションのそれと非常に類似しており、提案手法の有効性が確認された。

第 6 章では、マルチモーダル会話インタフェースのためのコンテンツ作成支援技術となるマルチモーダル生成機構 (Generation Module)、特に音声と同期した会話エージェントによるジェスチャーの自動生成について述べる。

ジェスチャーは発話中の重要な概念を強調するために用いられると言われている。そこで、プレゼンテーションのビデオデータを分析し、発話中の語彙・統語的情報とジェスチャー出現との関係を調べた。次に、この分析結果をジェスチャー付与ルールとして利用し、テキスト中の言語情報を用いてエージェント動作を自動的に決定、生成する機構、CAST を実装した。

次に、CAST をプレゼンテーション自動生成システム、SPOC に組み込み、利用者が好みの映像や画像のファイルを指定し、それへの説明文を入力するだけで、説明者キャラクターを登場させたテレビ番組風のメディアが自動的に生成される機構を実現した。

最後に、評価実験を行い、5 分程度の利用方法の説明で、全ての被験者が困難を感じることなくシステムの使い方を習得できること、また、SPOC により情報をわかりやすく伝えることができると感じていること等が明らかになった。これらの実験結果から、本システムにより、一般ユーザが手軽に会話エージェントによるコンテンツを作成できることが実

証された。

第 7 章では、将来課題について議論し、最後に第 8 章で本研究の結論を述べる。今後、マルチモーダル会話インタフェースが教育やネットワークコミュニティー等の実践的な場でより広く利用されるようになるにつれ、人間同士のインタラクションプロトコルの解明と、このプロトコルが対人工物のコミュニケーション様式をどのように規定しているのかの両側面について研究を進めることがよりいっそう重要になるであろう。現在の技術では人工物とのインタラクションは人間同士の対面コミュニケーションからは程遠いと言わざるを得ないが、本研究では学際的な研究アプローチを提案、実践することにより、対コンピュータ間コミュニケーションの研究領域において新しい方向性を示した。