

非言語情報を用いたマルチモーダル対話インタフェースの試作

向井理朗，関進，中沢正幸，綿貫啓子，三吉秀夫

技術研究組合 新情報処理開発機構 マルチモーダル機能シャープ研究室

シャープ（株）技術本部システム開発センター内

〒261-8520 千葉県千葉市美浜区中瀬 1-9-2 シャープ幕張ビル 19 階

TEL:043-299-8712 FAX:043-299-8709

E-Mail:{mukai, seki, nakazawa, watanuki, miyoshi}@iml.mkhar.sharp.co.jp

Abstract

本稿では対話時にユーザが発する音声から非言語情報を抽出し，相槌応答を行う対話システムについて報告する。人間同士の対話では非言語情報の占める割合が言語情報のそれよりもはるかに高いことが知られている。特に相槌には「聞いている」「同意する」など対話を円滑に進める役割がある。本稿ではユーザの発する非言語情報を基にして応答する対話システムについて報告する。

キーワード マルチモーダル，対話インタフェース，相槌，非言語情報

1. はじめに

近年情報家電機器が出現し，その操作が複雑になりつつある。リモコンや従来の GUI などのインタフェースでこれらの機器を操作するには限界がある。そこで我々は擬人化エージェントとの自然言語による対話の中から必要な情報を取り出し，機器の制御を行う対話型インタフェースの実現を目指している。

マルチモーダル情報を用いた対話システムとしては，すでに電総研の事情通口ボット[4]，RWCP マルチモーダル機能つくば研究室のマルチモーダルパソコン(MMPC)[5]等が提案されている。また相槌応答を行う対話システムには NTT の DUG-1 などがある。しかし，認識率の高い音声認識やジェスチャ認識の組み合わせだけでは，ユーザはシステムと対話してくれるとは限らない。擬人化エージェントの間にも人間同士のような快適な対話の実現されねばならない。そのためには発話交代や応答のタイミングなどの「間」が重要な役割を果たしており[2]，システムがユーザに合わせて自律的に応答することが必要である。我々はすでにマルチモーダル対話データベースの分析に基づいて「力学系対話処理モデル」を用いた自然で円滑な応答システムを提案している[1]。本稿では相槌の発生メカニズムに「力学系対話

処理モデル」を用いたマルチモーダル対話システム MAICO (Multimodal Agent Interface for Communication) について，その概要を述べる。

2. 対話処理モデル

我々は自律的に応答タイミングを生成する対話処理モデル[1]を提案した。ここでは「相槌をうちたい」という気持ちに対して力学系対話処理モデルを適用する。ある時刻 t におけるシステムの状態を x ，ユーザから得られる状態を y とするとき， f_{int} をシステム自身の状態， f_{ext} をユーザの影響とみなし，システムの状態の変化 q を(1)(2)式で表すものである。ここで a は気持ちの変化の速さを決める定数， b ， c はその強度を決める正の定数である。

$$\frac{dx}{dt} = q \quad (1)$$

$$\frac{dq}{dt} = f_{int}(x) + f_{ext}(y) \quad (2)$$

$$f_{int}(x) = -x^2 \quad (3)$$

$$f_{ext}(y) = \begin{cases} b(y-a)^2 + c & y \geq 0 \\ -b(y+a)^2 - c & otherwise \end{cases} \quad (4)$$

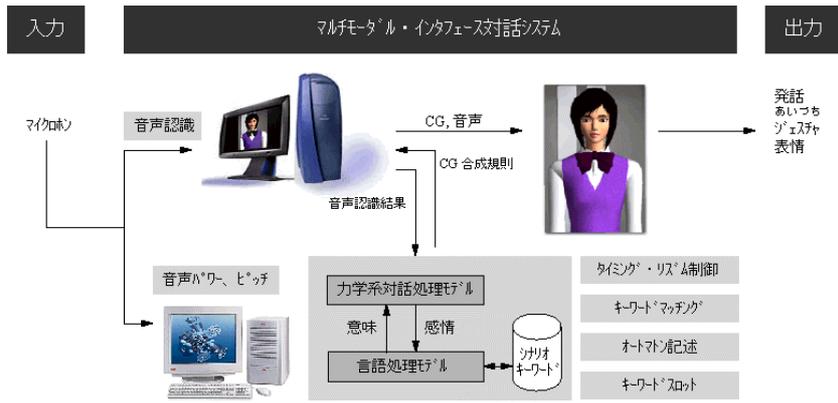


図 1：システム構成



図 2：対話風景

(2)式は自分のリズムとユーザの発声の強さなどの外部刺激により気持ちが変化することを示している。即ち、ユーザに対する反射的な反応しか示さない受動的なシステムではなく、自律的なリズムを持ち、且つ、ユーザの気持ちの変化に応じた反応を示すことができる。また、声の高さや大きさ、身ぶりの種類や大きさなどの複数の非言語情報を利用して、ユーザに応じた自然で連続的に応答するシステムへの拡張が可能である。

3. システム構成

スムーズな対話を実現するために、言語処理（対話モデル）、非言語処理（力学系対話処理モデル）、音声出力、CG 合成などの機能が必要である（図 1 参照）。我々はワークステーションと PC の 2 台を用いて対話システムを試作した。このシステムは力学系対話処理モデルに基づいて音声のパワー/ピッチを検出し、相槌のタイミング、音声出力の際のタイミングを生成する。

市販の音声認識ソフトを用いて音声認識を行い、その結果に対して簡単な意味処理を行う。対話モデルは認識結果に応じた CG の表情、身ぶり、音声出力のタイミングの調整を行う。

システムは言語処理に基づく応対をするだけでなく、力学系対話処理モデルに基づいて、相槌を打つ。人間が相槌を打つタイミングがいつも違うのと同様に、エージェントが相槌を打つタイミングもその都度異なる。相槌を打ちたいという気持ちが強かった時点がちょうどよいタイミングとなる。気持ちの強度によって、CG が頷くだけであったり、「はい」と相槌を打つなどの種類も複数存在する。

4. タスク

ユーザが自由に話す内容に対し、相槌をうちながら聞き役に徹する自由対話モードと特定タスクを遂行するタスクモード（「ファーストフードの応対」と「展示ホールの紹介」）の 2 つについてインプリメントを行った。CG エージェントに対し、「ハンバーガーを 2 つください」「飲み物には何がありますか？」などの対話からファーストフードにおけるオー

ダーの際の会話を実現した。また「展示ホールの紹介」ではユーザが展示ホール内のブースについて質問するとそれに応じた応答を行う。

自由対話モードではユーザの自由な発話に対し、相槌をうちながら聞き役を演じる。あらかじめ決められたキーワードに対して特定の応答を返すこととした。

数名の被験者について対話の様子を観察した。ファーストフードの応対ではユーザの発話が単語の羅列のように短い場合があり、相槌が有効に機能していない場合も見受けられた。また展示ホールの紹介ではエージェントが一方向的にしゃべりすぎる部分（紹介したい内容が多いため）もあるが概ねスムーズに対話が進行していた。

5. まとめ

本稿では力学系対話処理モデルに基づいたマルチモーダル対話システム MAICO について、その概要を報告した。本システムは相槌に着目し、力学系対話処理モデルを対話システムに適用した。これによりユーザの発話に対して、自律的に相槌をうち、対話をスムーズに行うことができる。

今後は実用化に向けて全機能を 1 台のパソコンへ実装するとともに、情報家電機器の制御やインターネット検索などのタスクでの検証を行っていく。

[参考文献]

[1] 関, 他: "力学系対話処理モデルに基づくリズムのある対話インタフェース", 情報処理学会 HI 研究会, 98-HI-80, pp.39-40, 1998
 [2] 綿貫, 他: "相槌位置の考察", 音響学会平成 10 年度春季講演論文集, pp.111-112, 1998
 [3] 中野 他: "柔軟な話者交代を行う音声対話システム DUG-1", 言語処理学会第 5 回年次大会論文集, 1999.
 [4] T.Matsui et al.: "An Office Conversant Mobile Robot That Learns by Navigation and Conversation", Proceedings of Real world computing Symposium, pp.59-62, 1997
 [5] 向井, 他: "マルチモーダルインタフェースのノートパソコンへの実装", 電子情報通信学会 PRMU 研究会 Vol.98, No.79, pp.69-75, 1998