

5E-3

力学系対話処理モデルを用いた 相槌システム

関進、綿貫啓子、中沢正幸、向井理朗、三吉秀夫

RWCP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

1. はじめに

我々は人間同士の円滑な対話を手本として、対話型のヒューマンインタフェースの構築を目指している。人間同士の円滑な対話では、如何にスムーズに情報の交換を行うかということだけでなく情報交換の過程、すなわち対話そのものが快適なものとなっている。対話の快適さ、気持ち良さには間や相の手のタイミング、テンポの良い発話交代など時間が重要な役割を果たしている。このような時間情報の伝達には、音声パワー・ピッチ、体の動きなど非言語的な情報が重要な働きをしていると考えられる。本稿ではマルチモーダル情報を取り扱うための「力学系対話処理モデル」[1]について説明し、それを相槌に適用した場合について述べる。

2. 力学系対話処理モデル

対話では言語情報・非言語情報が密接に絡んで進められるが、両者を統一的に取り扱うのは難しく、また、非言語部分のみでも対話の調整機能があるのではないかと考えられる。そこで、言語情報を切り離れたマルチモーダル情報を取り扱う対話モデルを提案し、その有効性や限界を見極めながら言語と統合していく方針を採る。

マルチモーダル情報を取り扱う場合の対話において、単に事象の発生順序だけでなく、発話速度や話者間・音声とジェスチャーのようなモダリティ間の同時性など実世界での時間情報が重要である。時間情報をモデルに明確に導入するために我々は力学系対話処理モデルを提案してきている。これはシステムが複数の内部パラメータを持ち、その遷移の仕方を時間による微分方程式などで記述する。すなわちシステム側の広い意味での気持ちを N 個の内部パラメータ q で表現し、その時間変化が自分自身の状態 q と M 個の外部パラメータ u （音声パワーなどのユー

ザ情報や、ユーザやシステムのおかれている環境など）によって定まる。

$$\frac{dq}{dt} = f(q, u) \quad (1)$$

モデルをこの様に記述することで実際の時間による振る舞いを記述することができるとともに、以下のようなメリットがあると期待している。

- 自律性

内部パラメータ q の遷移は、外部からの影響が無い状況、すなわち外部パラメータ u が0或いは定数であるときも、微分方程式に従って常に行っているため、単なるユーザに対する反射的な受け身の対話システムではなく、自分自身で対話に積極的に関わる能動的なシステムの構築が可能である。

- コンテキスト依存

外部からの入力がモデルに直接反映されるのではなく、内部状態 q によってその変化の仕方が異なる。この q の値は f の積分、すなわち自分自身も含めた過去の状態の積み重ね方によって異なるため、対話のコンテキストを反映させることができる。

- マルチモーダル情報の取り扱い易さ

マルチモーダルな非言語情報では、声の強さ、体の振りの大きさなどによって相手に与える効果も異なってくる。力学系対話処理モデルでは、ユーザの振る舞いで連続的に表す方が自然なものについて、そのまま、状態に反映することができる。また、逆にユーザへの出力にも連続的な量 q をもとにして、いろいろな強度で出力することが可能である。

Backchannel System using Dynamical Dialogue System

Susumu Seki, Keiko Watanuki, Masayuki Nakazawa, Toshirou Mukai, Hideo Miyoshi

Multimodal Functions Sharp Lab., Real World Computing Partnership in Sharp Co.

1-9-2 Nakase, Mihama, Chiba, Chiba 261-8520, Japan

3. 相槌モデル

3.1. 相槌のリズムの解析

我々が先に実施した実験では、人間が非言語情報を利用して相槌を打つことができることが示された[2]。この実験では、あらかじめ収録した語りかける口調の160秒の音声から、言語情報を落としたピッチのみの音声を作成し被験者に提示した。そして被験者が相槌している様子を録画し解析を行った。被験者の様子は1フレーム=1/30sec間隔で手作業で、「えー」「へー」などの音声や頷きの動作によって相槌を出している区間を $x[t]=1$ 、それ以外を $x[t]=0$ とラベル付けを行っている。この $x[t]$ の自己相関 $c[\tau]$ を求めた。その結果いくつかのデータにはピークが見られ(図1)、何らかの周期性があるものと推測される。

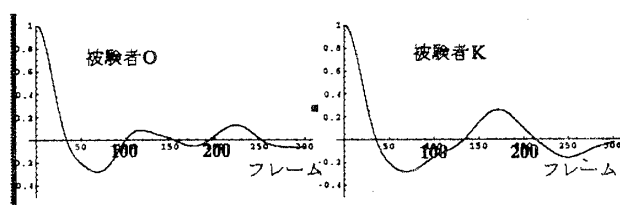


図1 被験者の相槌の自己相関

また、そのピークは入力ピッチの自己相関(図2左)とも被験者によっても異なることから、単に受動的に発話者のピッチに合わせた相槌ではなく、自分自身の状態に基づいて能動的、自律的に発話に関わるものがあると考えられる。

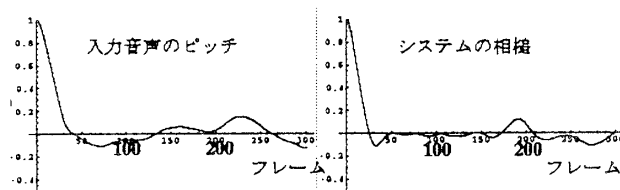


図2 入力音声のピッチ(左)とシステムの相槌(右)の自己相関

3.2. 相槌モデル

相槌そのものが本質的に持っている発話者に対する受動性と、データの解析から推測される自律性を兼ね備えたモデルとして我々は以下のように力学系対話処理モデルを相槌に適用した。

- ① 話者が話し続けると相槌を打ちたい気持ちが上昇、発話の途中でも気持ちが強くなれば相槌を打ち、同時に気持ちも下降する。
- ② 話者が話をやめると、相槌を打ちやすくなる。
- ③ 話者の話と聴取者の気持ちの盛り上がり具合が合えば、大きく影響を受ける。

相槌を打ちたい気持ち x とその変化 v をエージェントの状態とし、ユーザからの入力としてユーザの音声ピッチ V を用いて定式化した。

$$\frac{dx}{dt} = v \quad (2)$$

$$\frac{dv}{dt} = f_{\text{int}}(x, v) + f_{\text{ext}}(x, v, V) \quad (3)$$

$$f_{\text{int}}(x) = -\omega^2 x \quad (4)$$

$$f_{\text{ext}}(x, v, V) = \begin{cases} b(V-v-a)^2 + c \cdots & \text{if } (V \geq v) \\ -b(V-v+a)^2 - c \cdots & \text{otherwise} \end{cases} \quad (5)$$

上で述べた相槌モデルに基づいて、実時間相槌システムを構築した。これはマイクからユーザの音声を入力し、CGエージェントが頷きや「えー」「はい」などの相槌を出力するものである。3.1の相槌実験で用いた音声入力に対してシステムに相槌を打たせた。このとき ω などのパラメータの設定によりシステムのユーザからの影響の受けやすさや自律的な周期等個性を出すことが出来る。相槌の自己相関の一例を図2右に示す。

4. おわりに

今後は、マルチモーダル情報を提示された場合の人間の相槌についてその関係の詳細を調べ、より適切な振る舞いを表現するシステムの構築を目指す。また、言語と非言語情報を統一的に取り扱う対話システムの構築を目指し、簡単な言語情報を取り扱うことが出来るシステムを作る予定である。

参考文献

1. 関, 木山, 日出, 綿貫, 荒巻. "力学系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, 関, 木山, 荒巻. "あいづち位置の考察", 音響学会平成10年度春季講演論文集, pp. 111-112, 1998.