

力学系モデルに基づくリズムのある ヒューマンインタフェース

関進、木山次郎、日出晴夫、綿貫啓子、荒巻隆志

RWCPマルチモーダル機能シャープ研究室 シャープ(株)AVC研究所内

1. はじめに

気持ちの良い円滑な対話においては独自のリズムやテンポがあるのではないかとされているが、我々はそのリズム生成には相槌が重要な役割を果たしていると考えている。相槌に関する研究では例えば[1]等があるが基本的に、これらは発話に対する受動的な反応として相槌を捉えている。発話があって相槌が打たれることからこの捉え方は当然であるが、我々はこれ以外にも、聴取者の対話への積極的関与として相槌を能動的に捉えることが可能ではないかと考えている。そこで我々は相槌モデルを提案し、これを用いた相槌を打つシステムを試作した。本稿では、その概要について述べる。

2. 相槌のリズムの解析

我々は相槌にもリズムがあるのではないかと考えあらかじめ収録した語りかける口調の160秒の音声に対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間隔で手作業でラベル付けを行った。時刻 t 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 $x[t]=0$ とし、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、何らかの周期性があるものと推測された。また、その周期が被験者によって異なることから、この周期性すなわちリズムは単に受動的に発話者のキーワード、抑揚、発話ポーズに合わせた相槌だけではなく、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。我々はこの自律的リズムにより、乗りの良い会話が形成されるのではないかと考えている。

3. 相槌モデル

解析結果が示すように、人が相槌を打つ場合単純に発話者に反応しているのではない可能性がある。そこで、話者の発話による外因と聴取者の自律リズムの調和から相槌が生まれるものと考え、これを実現するために力学系対話処理モデルを用いた。

3.1. 力学系対話処理モデル

力学系対話処理モデルではシステムの状態の遷移の仕方を微分方程式や差分方程式で記述する。このモデルには以下のようなメリットがある。

- 自律性

内部状態の遷移を微分方程式に従って常に行っているので、単なる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

外部からの入力があるまま状態に反映されるのではなく、状態によって同じ入力でもその変化の仕方が異なる。そのため、過去の状態の積み重ね方が現在の状態へ大きく影響を与えるので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

対話においては言語情報のみならず、非言語情報が重要な役割を果たしている。非言語情報のやり取りでは情報の内容と共に、そのやり取りされた時刻が重要である。またその取り扱いも意味付けされた記号的なものとしてよりも、時刻による変化など時系列パターンのまま取り扱う方が自然である。記号的な遷移で取り扱いをしようとする場合、多くのノードとその遷移ルールを記述しなければならず現実的ではない。力学系対話処理モデルでは内部状態を連続的なパラメータで表現し、その

遷移を簡単な方程式で記述できる。このとき、連続的なマルチモーダル情報を連続的なエージェントの内部状態に結び付けることが可能である。

3.2. 相槌への適用

相槌モデルを構成する方針として以下のような振る舞いの実現を目指す。

- ① 話者が話し続けると相槌を打ちたい気持ちが上昇し、発話の途中でも自律リズムで相槌を打ち、同時に気持ちも一旦下降する。
- ② 話者が話をやめてポーズになると、相槌を打つと同時に気持ちが急速に下降する。
- ③ 話者の話と聴取者の気持ちの盛り上がりがあえば、大きく気持ちが盛り上がり、大きな相槌を打つ準備を行う。

以上の方針に基づき、以下の方程式で振る舞いを表現する。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パワーの対数 V を用いる。

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

$$\frac{dv}{dt} = f_{int}(x, v) + f_{ext}(x, v, V) \quad (2)$$

(2)式の第1項は自律的な状態の遷移の効果を表わし、①を満たすものとして単純な調和振動子

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

を採用する。ここで ω は正の定数である。(2)式の第2項は外部、すなわちユーザからの影響を表わす項で③の効果を実現するために $y=eV-v$ の関数とする。ここで e は正の定数とする。そして、両者の気持ちの盛り上がり一致する $eV=v$ すなわち $y=0$ 近辺ではエージェントに影響を大きく与えるとともに②の効果をに入れるために次のような関数で表わす(図1)。

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

ここで、 a, b, c は正の定数であり ω と同様にエージェントの個性を表わす。

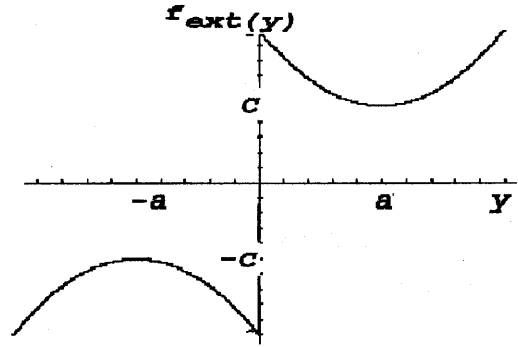


図1 $f_{ext}(y)$ の模式図

4. 実時間相槌システム

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。CGエージェントの音声は、あらかじめ人間が発話したものを録音して用いた。相槌を打つタイミングは x の極大になったときとし、 v の変化率によって相槌の強度を変えた。ここで相槌の強度は頷きの大きさとしゃべり方に反映させた。

使った感想としては、話の途中で相槌を出したり、ポーズで打つたりと、ある程度うまく相槌を打っており、タイミング良く打たれるとユーザは快適さを感じることができる。しかしながら、相槌を打って欲しいときに打たないこともあり、改良の余地もある。今後は人間のデータ解析を対話モデルと比較し、リズム以外の要素もモデルに反映していくとともに、このシステムを実際の人の対話で用いることでインタラクションとしての効果を調べていく。

5. 参考文献

1. Watanuki, K., Sakamoto, K. and Togawa, F. "Analysis of Multimodal Interaction Data in Human Communication", *Proceedings of ICSLP 94*, S17-8.1, pp. 899-902, 1994.