

視線制御モデルを用いた擬人化音声対話エージェントの提案

Gaze Modeling and Analysis for Spoken Dialog Systems with Anthropomorphic Agents

中沢 正幸 西本 卓也 嵯峨山 茂樹
Masayuki Nakazawa Takuya Nishimoto Shigeki Sagayama

東京大学大学院 情報理工学系研究科
Department of Information Physics and Computing, The University of Tokyo

For spoken dialog systems with anthropomorphic agents, it is important to give natural impressions and real presence to human. For this purpose, gaze controls of the agent which are consistent with the spoken dialogs are expected to be effective. Our approach is based on the following hypotheses: (1) An agent performs the dialog concurrently with the intentional controls of the gaze to retrieve the information and to give signals. (2) The movement of the head and eyeballs is based on mathematical models. In this paper, we propose the gaze model of gazing defined by Mass-Damper-Spring Model.

1. はじめに

擬人化音声対話エージェントにおいて、対話の流れに応じて適切にアイコンタクトを行うことは、対話相手である人間に自然な印象を与え、擬人化音声対話エージェントの実在感を高めるうえで重要である。我々は、音声合成の分野において声帯振動機構に基づいたモデルが合理的な方法で多様な基本周波数 (F_0) パターンを説明することに着目し、さまざまな対話現象を力学現象に等価変換することで擬人化音声対話エージェントの心的状態と挙動を結びつけるモデルの構築を目指している。具体的には、

1. 擬人化音声対話エージェントは相手に関する情報を得たり相手に合図を送ったりするために能動的に視線を動かしながら対話を行う
2. 擬人化音声対話エージェントの頭部及び眼球の動きは数理的な視線制御モデルに従う

という仮説に基づいて、対話の流れに応じて擬人化音声対話エージェントの視線の動きを制御する手法について検討している。

我々は、「擬人化音声対話エージェント基本ソフトウェア」[Sagayama 02] の基本通信プロトコルに適合し動作する、3D CG で構築された擬人化音声対話エージェント (図 1) を開発した。この 3D CG 擬人化音声対話エージェントは、全身を持ち、腕・手の動作、表情、唇の形状、視線・頭の動作をそれぞれ独立に制御することができる。我々は、この擬人化音声対話エージェントを用い、音声認識・画像認識等の言語によるコミュニケーション能力を加えることで、視線や表情、身振り・手振り、感情などの非言語情報を表現でき、より実在感のある人間型の音声対話環境を構築することを目指している。

この目的のために擬人化音声対話エージェントが実現すべきことは、人間と同じような動作を行い、自然な動きを実現することである。特に視線は、広い範囲で捉えると「表情」の一部であり、自然な動きの実現と密接な関係がある。

通常、話し手は最初に聞き手と視線を合わせてから、対話を開始し、話を始めたら聞き手から目をそらす。話の途中で話し手が聞き手を見ることで、聞き手が自分の話をよく聞いて

いるかどうか、あるいは理解しているかどうかを確認できる。また、視線を相手の目に止め、相手に発話権を渡したいという合図を送ることもできる。擬人化音声対話エージェントにおいてこれらの視線制御を行うことは、人間同士の対面対話に慣れている人間にとっては機械と対話しているという違和感の解消につながるため、より話しかけやすい対話システムを実現するうえで重要である。

視線が会話のやり取りを調整することは、これまでの研究の中で、Duncan & Fiske [Taibo 98, Duncan 77] らによっても明らかにされ、前田ら [Maeda 03] や深山ら [Fukayama 02] の追実験によっても確かめられている。特に、視線制御において、Lee ら [Lee 02] は、対話における人間の眼球運動を観察し、凝視・非凝視時間を 2 次の多項式により表し、凝視位置の分布を用いてランダムに凝視動作をさせるモデルを提案している。また、Bilvi ら [Bilvi 03] は、信念ネットワークにより、各時刻における直前の話者・非話者の凝視状態、話者・非話者の動作状態、発話・非発話状態から話者の次の凝視が発生する確率を推定している。しかしながら、これらの研究では、凝視・非凝視のタイミングと凝視量・非凝視量をモデル化することを主眼としており、その凝視・非凝視位置は、分布を用いてランダムに生成している。

これらの研究のように、擬人化音声対話エージェントの視線を制御する場合には、個々の挙動に対応する頭部や眼球などの座標移動コマンドをあらかじめ用意しておき、対話の流れに応じてコマンドを実行する、といった方法がとられている。

しかし、前述したような挙動を擬人化音声対話エージェントを用いて実装する場合には、視線を相手に向けるだけでなく、視線をそらしていく場合にも自然な動作を実現するために、何らかの統一的な視線運動モデルが必要になる。つまり、「相手の顔を凝視する」「視線をそらす」などの行為が、座標移動のコマンドによってもたらされるのではなく、状態の変化に伴うモデルのパラメータの更新に対応しており、実際の頭部や眼球などの座標はモデルから逐次的に計算される、と考えることが望ましい。

また、擬人化音声対話エージェントの外見や声に個性をあたえることが可能となってきたが、それに伴った動作において個性が存在しないのは、人間に不自然さを感じさせる恐れがある。我々は、視線制御モデルに、振る舞いにおける個性 (個人差) をパラメータとして含めることを目指しており、これにより、外見や声と同時に挙動にも個性をあたえることを実現し

連絡先: 中沢 正幸, 東京大学大学院 情報理工学系研究科, 〒113-8656 東京都文京区本郷 7-3-1, TEL/FAX: 03-5841-6902, nakazawa@hil.t.u-tokyo.ac.jp

たいと考えている。

このようなモデルを用いることによって、擬人化音声対話エージェントにあたかも現実の人間のように振る舞わせると同時に、擬人化音声対話エージェントと対話する人間に、擬人化音声対話エージェントが意識を持ち能動的に何かに興味を持っているかのように感じさせることが、本研究の目標である。

本報告では、擬人化音声対話エージェントの二次遅れ系標準形を用いた視線運動に関する基本的な定式化の概要について述べる。



図 1: 3D CG 擬人化音声対話エージェント (顔と全身)

2. 音声対話に伴う視線制御

2.1 二次遅れ系標準形

我々は、人間型の擬人化音声対話エージェントの眼球及び、頭部(顔方向)の運動自由度を考慮した、数理モデルによる視線運動の自動生成を目的としており、スクリーン座標系の平面上的2軸(x軸, y軸)における点として視点を定義し、視線運動モデルの予備的な検討を行なった[Nakazawa 04]。さらに本報告では、対話における話者の状態や意図などを力学現象に置き換えることを目指し、伝達関数表現で式(1)

$$\frac{k\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (1)$$

のような二次遅れ系標準形(分母のsが2次式、固有振動数 ω_n 、減衰係数 ζ 、ゲイン係数k)をモデルに導入し、頭部運動を伴う視線運動モデルの提案を行う。

二次遅れ系はステップ入力への応答に際して、 $\zeta = 1.0$ (臨界制動)を境界として振動現象の有無が変わる性質を持ち、質量、粘性摩擦、バネを持つ機械系の基本的な制御モデルとされる。

この制御モデルを視線の運動に対応させるにあたっては、感情の変化、うなづき、発話などが特定のステップ入力などに対

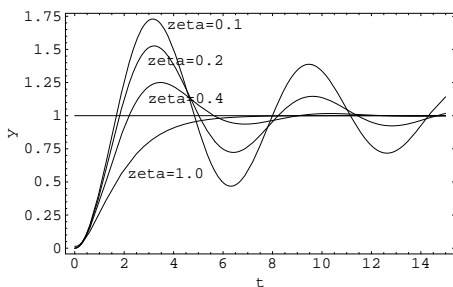


図 2: ステップ応答に対する時間応答 ($\zeta = 0.1, 0.2, 0.4, 1.0$)

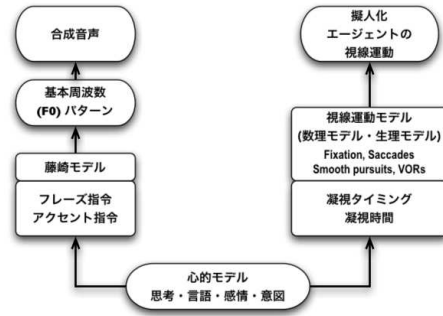


図 3: 心的モデルを基本とする合成音声と擬人化音声対話エージェントの視線運動の関係

応したり、モデルの質量やバネ係数を変化させる、といった対応付けを行うことが必要となる。

自然な視線運動を単純な制御モデルで実現することが本研究の目的であり、得られる制御モデルはある場合は物理的に人間の身体運動に対応し、ある場合は心理現象を力学系に等価変換するためのモデルとして有用であると期待される。

2.2 視線運動モデルの定式化

我々は、予備実験として視線運動モデルを一次式で表現してきた[Nakazawa 04]。しかし、その場合頭部を含んだ視線の運動は、等速度となり、実際の様々な加速度分布をもった運動を表現できていなかった。そのため、二次遅れ系標準形をモデルに用いることで、より多様な運動を表現することを試みる。本モデルでは、次のような仮定を立てた。

- システムには情報欲求度があり、この欲求度が高まると相手の方に視線(頭部)を向ける。視線を向けた時間に応じて、この情報欲求度は降下する。
- システムには凝視欲求度があり、この欲求度が高まると相手の方に視線(眼球)を向ける。視線を向けた時間に応じて、この凝視欲求度は降下する。凝視欲求度は、凝視位置、凝視時間をパラメータとする。

視線運動 $E(t)$ を以下のように定義する。

$$E(t) = E_{base} + E_{head} + E_{eye} \quad (2)$$

ここで、 E_{head} は、頭部運動を表し、 E_{eye} は、眼球運動を表す。それぞれの運動は、二次遅れ系標準形(1)で表現されるとする。この時、 E_{head} には、システムが持つ情報欲求度がステップ入力(δ 関数)として与えられる。 E_{eye} には、システムが持つ凝視欲求度がインパルス入力として与えられる。なお、 E_{base} は、システムが定常的に向けている視線方向(頭部及び眼球)の分布である。頭部及び眼球運動は、(2)により生成される。この時、眼球は、物理的にモーメントが小さいと考えられ、頭部は逆に大きいと予想される。この眼球と頭部のモーメントの違いが、固有振動数 ω_n 、減衰係数 ζ 、ゲイン係数kに表れてくると考えている。

3. まとめ

本報告では、擬人化音声対話エージェントが表現可能な非言語情報の一つである視線に着目し、視線制御モデルの基本的な定式化を行った。また、視線運動データの分析を行い、特に相手の顔を凝視している状態における検討を行った。

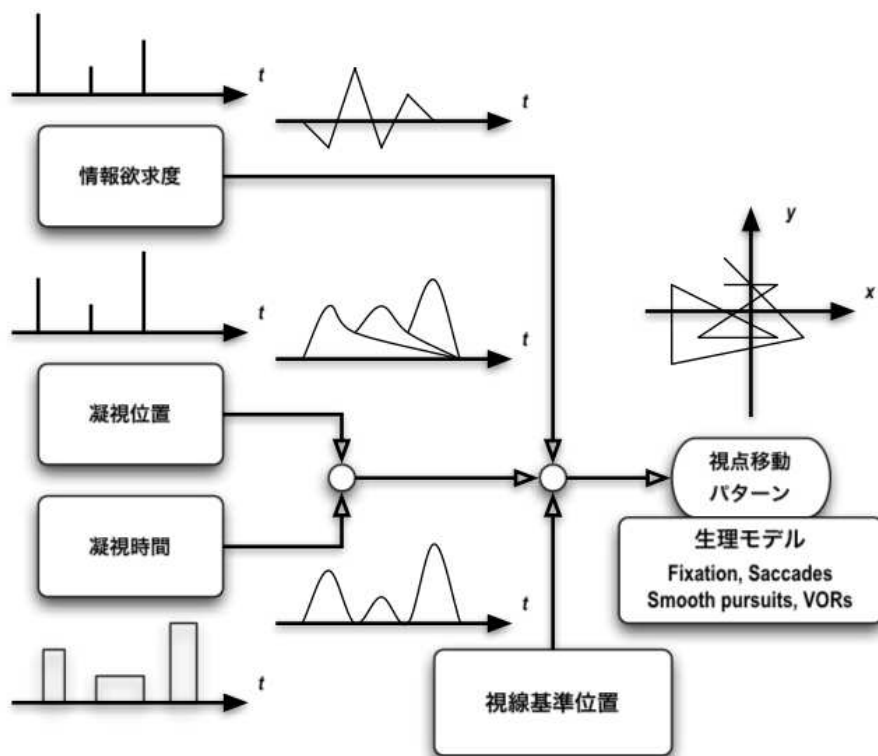


図 4: 凝視位置, 凝視時間, 情報欲求度, 視線基準位置に基づく視線制御モデルのブロック図

我々の提案は、擬人化音声対話エージェントにおいても人間と同様に、視覚情報が受動的に「見えているもの」ではなく、能動的に「見るもの」[Gibson 79, ?] であるという前提に基づいている。擬人化音声対話エージェントを用いて、たとえ模倣的にであっても、外界の様子を探ったり、相手の表情を読みとったりしながら活動に必要な情報を取得していく様子を表現することで、より自然な人間と擬人化音声対話エージェントとの対話を実現できると考えている。また、実際に視覚情報を利用して対話を行う場合にも、振る舞いが人間らしく見えることを考慮して擬人化音声対話エージェントの見掛け上の視線を制御することが求められる。

今後は、制御モデルを視線の運動に対応させるための感情の変化、うなづき、発話などが特定のステップ入力などに対応したり、モデルの質量やバネ係数を変化させる、といった対応付け、凝視位置、凝視時間の対応付けを行い、モデルの詳細化および擬人化音声対話エージェントへの実装・評価を行っていく予定である。

謝辞

本研究の一部は東京大学 21 世紀 COE プログラム「情報科学技術戦略コア」(実世界情報システムプロジェクト) の支援を受けた。また、本研究は、RWC 研究用マルチモーダル対話データベースを使用した。

参考文献

[Bilvi 03] Massimo Bilvi, and Catherine Pelachaud: Communicative and Statistical Eye Gaze Predictions, *In Proc. of Embodied Conversational Characters as Individuals, AAMAS*, (2003).

[Duncan 77] Duncan, S.D., Jr., and Fiske, D.W.: Face-to-face interaction: Research, Methods, and Theory, Hillsdale, N.J: Lawrence Erlbaum, (1977).

[Fukayama 02] 深山篤, 大野健彦, 武川直樹, 澤木美奈子, 荻田紀博: 擬人化エージェントの印象操作のための視線制御方法, *情処論誌*, Vol. 43 No. 12, pp. 3596-3606, (2002).

[Gibson 79] James J. Gibson: *The Ecological Approach to Visual Perception*, (1979) (邦訳: 古崎敏, 古崎愛子, 辻敬一郎, 村瀬旻 共訳: 生態学的視覚論 -人の知覚世界を探る-, サイエンス社, (1985)).

[Ikeda 75] 池田光男: 視覚の心理物理学, 森北出版, (1975).

[Ikeda 88] 池田光男: 眼はなにを見ているか -視覚系の情報処理-, 平凡社, (1988).

[Lee 02] Sooha Park Lee, Jeremy B. Badler, and Norman I. Badler: Eyes Alive, *In Proc. of ACM SIGGRAPH*, (2002).

[Maeda 03] 前田真季子, 堀内靖雄, 市川薫: 自然対話におけるジェスチャーの相互的關係の分析, *情処研報, IPSJ-SIGHI-102-7*, pp. 39-46, (2003).

[Nakazawa 04] 中沢正幸, 西本卓也, 嵯峨山茂樹: 擬人化音声対話エージェントにおける視線制御モデルの提案, *人工知能学会 SIG-SLUD-A303*, pp. 21-26, Mar, (2004).

[Sagayama 02] 嵯峨山茂樹, 他: 擬人化音声対話エージェントツールキット Galatea, *情処研報, IPSJ-SLP-45-10*, pp. 57-64, (2002).

[Taibo 98] 大坊郁夫: しぐさのコミュニケーション -人は親しみをどう伝えあうか-, サイエンス社, (1998).