

視線制御モデルによる擬人化音声対話エージェントの制御

Gaze Modeling 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 an experiment to observe the motion of the head and gazing for modelling effective behavior of an agent.

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な情報の授受の制御を行うことは、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、様々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、次のような仮説に基づいて、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。



図 1: 3D CG 擬人化音声対話エージェント usherette (顔と全身)

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

我々は、擬人化音声対話エージェントの実在感を検証するプラットフォームとして、「擬人化音声対話エージェント基本ソフトウェア」[嵯峨山 02] の基本通信プロトコルに適合し、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 3D CG で構築された擬人化音声対話エージェント usherette (図 1) を開発した。この擬人化音声対話エージェントに音声認識・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的な音声対話環境を構築することを目指している。擬人化エージェントの振る舞いに個性 (個人差) をパラメータとして含めることによって、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ことが本研究の目標である。

本報告では、アイコンタクトにおける視線運動・頭部運動を二次遅れ系標準形を用いて、パラメトリックにモデル化することを目的に、アイコンタクトが有効に機能 (情報の授受の制御が行われる) している状況の観察・分析を行うための予備実験について述べる。

2. コミュニケーションにおける視線の役割

2.1 対話と視線行動

Argyle[Argyle 88, 大坊 98] によると、視線という概念の中には、相手に向けた視線の時間、相互視 (相手の顔の一部への視線)、話しをしながら相手を見る、相手の発言を聞きながら見る、2~3 秒の瞬間的な視線 1 秒程度の相互の瞥視視線がどこに長く向けられるかという固着 (視線のたどる軌跡)、瞳孔の大きさ、目の表情 (白目の割合で相手を睨む、射るような視線)、相手を見ていないときはどこに視線を向けているか、瞬目 (まばたき) の割合があるとしている。

Kendon[Kendon 67, 大坊 98] は、視線活動は相手に注意を向け、伝達する用意のあることを相手に知らせる動きを持っていることを指摘している。さらに、1) 感情表出を中心とする対人的態度の表出、2) 自分の働きかけに対するフィードバックを求めることを含む情報収集、3) 会話の流れを調整する機能を挙げている。1) の対人的態度の表出機能は、対人関係における接近一回避の感情を表現するもので、他者への関与、親和性、好意を示す働きによるものとしている。関心や好意を持つものに対しては、向ける視線が多くなることはこれまで繰り返され確認されている。

2.2 対話中の視線の役割

Kendon[Kendon 67, 福井 84] は、対話中の視線行動を 3 つの機能に分類している。

1. モニター (monitoring) 機能。発語を終えるべきか続けるべきかを、聞き手が凝視しているかどうかによって確かめるモニターとして使っている。

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

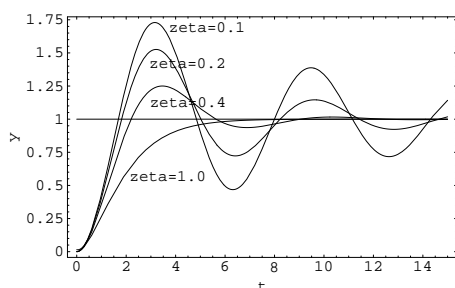


図2: ステップ入力に対する時間応答 ($\zeta = 0.1, 0.2, 0.4, 1.0$)

- 調整 (regulation) 機能. 会話が好ましいものであるかどうかを相手の目の動きによって察知して, 相手が目を輝かせて見つめるような話の内容へと調整していく.
- 無言の表出 (expressive) 機能. 会話が効果を及ぼしているか, あるいはマイナスの効果を与えているかを相手に知らず働きをしている.

目の動きは会話の内容や会話の進行具合と大いに関わりがある. 流暢に話が進んでいるときは話しては相手をあまり見ない. しかし, 会話が滞りがちの時は, 話しては相手を頻繁に見る. 話し終わりと話し始めとのつなぎの休止期間で, 話し手の相手を見つめる回数は, その休止期間が息継ぎとしての話しの区切りの時は相手を見る場合が多いが, とぎれがちの会話では, その休止期間が話すことのため生じているときは相手をあまり見ない.

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.

3. 音声対話に伴う視線制御

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の 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ってきた [中沢 04]. さらに, 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, 頭部運動に伴う視線運動モデルの提案を行ってきた.

二次遅れ系はステップ入力 (図 2) 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.

4. 視線運動・頭部運動データ収集の予備実験

4.1 実験の目的

情報の授受の制御がアイコンタクトにより有効に行われるような状況, 例えば, 聞き返しにより質問内容を正確に理解した上で対話が円滑に進行するといった状況を観察・分析し, 視線運動・頭部運動をパラメトリックにモデル化するための基礎データの収集を目的とする.

今回の予備実験においては, 頭部運動を 3D モーションセンサ (セラミックジャイロ, 加速度センサ, 地磁気センサを組み合わせた姿勢角度を検出するセンサユニット) を利用して, 頭部運動の分析を行っている.

ユーザは, 自然な視線・頭部運動を行うことで, 自然で円滑な対話を行っているが, 何の制約も加えない自由な状況下でのデータ収集は, 視線・頭部運動以外の他の様々な要因が影響を及ぼし, 視線・頭部運動のパラメトリックなモデル化のためには適さない. そのために, ユーザの反応に制限を加えた, 特定の条件下での視線・頭部運動データを収集する. パラメトリックな (藤崎モデル的) モデルにおけるアクセント成分・フレーズ指令に相当するものが明確であるような実験条件でデータを収集することにより, パラメータの推定が容易になると考える.

与える課題は, 計算タスクとする. 出題は音声で行い, 回答は非音声 (筆記) で行うこととし, 実験状態の自然性よりも統制を重視する. 音声で聞き返せないのは不自然であるが, 「音声による出題を聴取する」という現象そのものは自然である. 非言語情報がより明確に表現されるはずであり, 実験の目的に即していると考えている. 計算タスクを選ぶ理由として, 「問題を聞いて回答する」という, 1 つの単純な行為であること, 個人の知識経験に影響されにくい, 繰り返しても学習効果が少ないことが利点として挙げられる.

4.2 実験方法

実験は, 出題者と回答者との 2 者の対話において, 回答者が出題者に対して, 適切にアイコンタクトを行うことにより, 出題者が回答者の様子を視覚的に読み取って, 適切に問題を再発声するという課題を行う.

課題は計算問題 (簡単な複数術の暗算) とし, 出題時の声の大きさ・発話速度に変化をつけることにより, 出題内容が「聞き取りにくい, 理解できない」という状況を作る. 出題者はヘッドホンを装着して音楽等を聴取している状態, すなわち, 回答者の音声を聞き取ることができない状態にする. 一方, 回答者はヘッドホン等を装着せず, 自然な状態にする.

回答者が答えるときは, 回答の筆記のために用紙に視線を置き, 記入 (出題者はヘッドホンを装着しているため, 発声しても回答を聞くことができないため) してもらう. 回答が終わった後は, 再び出題者の方を見て, 次の問題に移ってもいいことを伝える. 回答者は, 出題者の声が小さくて聞き取れない場合や発話速度が速すぎて問題が理解できない場合には, 出題者に再度, 問題を発声してもらうよう依頼する.

このような状況下では, 出題者は, 回答者から受ける視覚情報のみを頼りに, 問題の再発声を行う. また, 回答者は, 出題者が視覚のみを頼りにしていることが事前に分かっているため, 音声だけではなく, 視線や頭部の仕草などの非言語情報を優位に使って, 出題者に情報を伝える. 以上のような実験環境下では, 出題者と回答者がそれぞれ, 発話者 (仕草等も含む), 聴取者になり得る. 発話者・聴取者という視点から整理した表出イベントは以下ようになる. 本実験では, 回答者の出力情報 (聴取行為及び発話行為) を観察する (図 3).

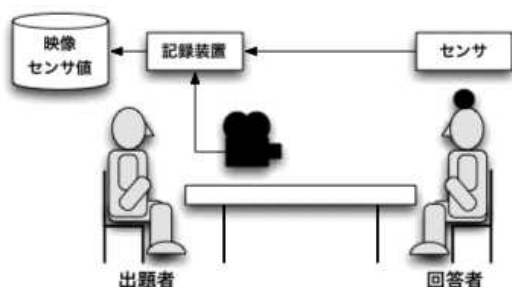


図 3: 実験環境 (カメラの映像とセンサ値とを同期を取りながら実験の様子を記録)

表 1: 出題者・回答者の行為と表出イベント

	出題者	回答者
発話行為	(a) 出題 (話し始め・終わりイベント)	(b) 聞き返し (聞き返し要求イベント), (d) 回答 (回答イベント)
聴取行為	(e) 回答受理 (回答受理イベント)	(c) 受理表明 (受理表明イベント)

- 発話行為としてのイベント：話し始め・話し終わり (聴取者の顔への注視), 回答 (聴取者の顔への注視), 聞き返し要求 (聴取者の顔への注視)
- 聴取行為としてのイベント：回答受理 (聴取者の顔への注視, 頷き), 受理表明 (頷き)

出題者・回答者の行為とそれぞれの表出イベントは表 1 (表中の記号 (a)~(e) は, 図 4, 5 の記号と同一) のようになる。

図 4 は, 聞き返しのない場合の対話の流れを示し, 図 5 は, 聞き返しのある場合の対話の流れを示す。それぞれの対話の具体的な流れは以下のように分析できる。

聞き返しのない対話の場合は, (a) 出題者から回答者に問題が出題される, (c) 回答者は問題が理解できたことを暗黙的に, もしくは明示的に出題者に伝え, 出題者はそのことを理解することで, 再出題の必要がないことを知る。(d) 出題者は, 回答者の回答が終了したことを理解し, (e) 次の問題に移ることを回答者に仕草等で伝え, 回答者に次の計算問題に対する回答の準備をさせる。

聞き返しのある対話の場合は, (a) 聞き返しのない対話の場合と同様に問題が出題者から回答者に問題が出題されるが, (b) 回答者は, 問題が聞き取れなかった, もしくは理解できなかったことを出題者に伝える。(a) 出題者は, 回答者の様子から再出題の必要があることを知り, 再度, 問題を出題を行う。(c) 今度は, 回答者は, 問題が理解できたことを暗黙的に, もしくは明示的に出題者に伝え, 出題者はそのことを理解することで, 再出題の必要がないことを知る。(d) 出題者は, 回答者の回答が終了したことを理解し, (e) 次の問題に移ることを回答者に仕草等で伝え, 回答者に次の計算問題に対する回答の準備をさせる。

4.3 記録方法

実験の様子は 1 台のビデオカメラと 1 台の 3D モーションセンサ (セラミックジャイロ, 加速度センサ, 地磁気センサを組み合わせた姿勢角度を検出するセンサユニット: NEC TOKIN 製 MDP-A3U9S) を用い, プログラムによりビデオ映像 (30fps) とセンサ値 (15fps) の取得とを同期させながら記録した。出力



図 4: 聞き返しのない場合の対話の流れ

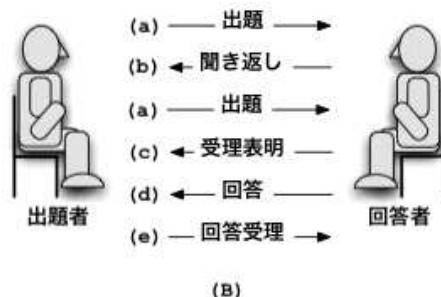


図 5: 聞き返しのある場合の対話の流れ

されるセンサ値は, 頭部の左右運動角度 (Yaw), 斜め運動角度 (Roll), 上下運動角度 (Pitch) を表す (角度単位は degree)。

4.4 手続き

予備実験として 1 名のデータのみの収録を行った。出題者と 1 名の回答者 (被験者) を用意し, 出題者及び回答者は, お互いに向き合って着席させ, 次のような指示を行った。

- 出題者への教示：これから 5 問の計算問題を問題を出題してもらいます。読み方の指示が書かれていますので, その通りに読んでください。回答者が問題を聞き取れていないように感じたら, すぐに問題を復唱してください。回答者が考えているようなら 10 秒以上, 答えるまで待ってください。回答者がその回答が終わったら, 次の問題に進んでください。
- 回答者への教示：これから 5 問の計算問題を問題を出題します。計算結果を暗算して, 答えを紙に書いてください。回答の時のみ, ペンを持って前に置かれている紙に書き, 回答が終わったらペンを置いてください。問題は, 声を出して復唱してもかまいませんが, 出題者は, ヘッドホンをしているため, 回答者の声を聞くことはできません。問題が聞き取れない場合は視線や表情, 仕草や声などで出題者に伝えてください。聞き返し以外は, できるだけ頭や目を動かさないでください。課題達成までの時間を計測しますので, なるべく早く達成してください。
- 計算問題例: 「^{イチゼロゼロゼロ}1000 たす ^{イチゼロゼロゼロ}1000 たす ^{イチゼロゼロイチ}1001 は?」, 「^{イチ}1 たす ^{イチゼロゼロゼロゼロゼロ}1000000 は?」

4.5 結果

収録した頭部運動情報 (Yaw, Pitch, Roll) を図 6 に示す。予備実験は, 5 問の計算問題を問題を出題し, 実際に 3 問目, 4 問目, 5 問目で聞き返しの要求が起こった。回答の際は, 机の上にある紙に記入してもらったため, 実際に頭部の上下運動が起こっており, 図 6 に 5 箇所の 40 度前後の大きな上下運動 (図 6 の最下部の折れ線グラフ) で示されている。この 40 度前後の頭部の大きな上下運動 (Pitch) をしている箇所は, ビデオ映像から回答者が回答を記入している時刻であることが確認された。

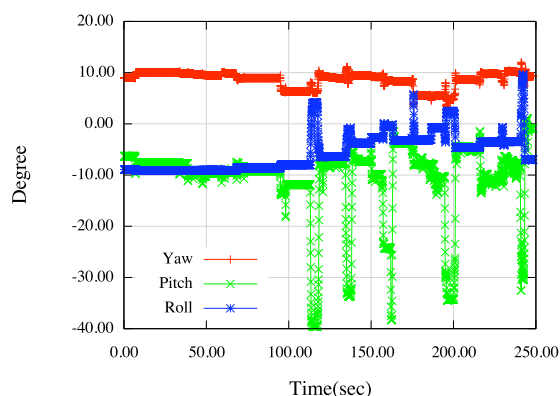


図 6: 頭部運動 (図中の折れ線グラフは上から, 左右運動 (Yaw), 斜め運動 (Roll), 上下運動 (Pitch) を示す)

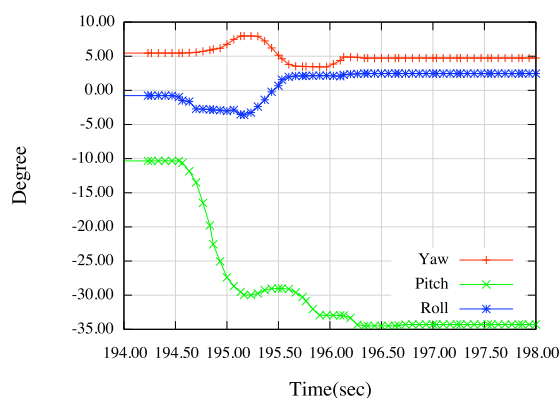


図 7: 4 問目の出題時の聞き返し動作 (195 秒前後) における頭部運動 (図中の折れ線グラフは上から, 左右運動 (Yaw), 斜め運動 (Roll), 上下運動 (Pitch) を示す)

また, 頭部の小さな上下運動 (Pitch) は, 頷きの箇所であることが同様にビデオの映像から確認された。

頭部の左右への回転運動 (Yaw) は, 大きな変化がない。しかし, 頭部を斜めにかしげるような運動 (分からない時など首をななめにひねるような動作: Roll) は, 回答の時刻及び, 3 問目, 4 問目, 5 問目の回答時刻前にも発生している。聞き返し要求が起った時は, 頭部の上下運動 (Pitch) と頭部の斜め運動 (Roll) が小さい運動を起こしていることが分かる。すなわち, この回答時刻前の Pitch と Roll の動きは, 聞き直しの際の頭部運動の特徴が現れていると考えられる。

4.6 考察

収集したデータからは, 頭部運動の詳細なデータ及び, (b) 聞き返し要求イベント (頭部の上下の小さな運動 (Pitch) と頭部の斜めの小さな運動 (Roll)), (c) 受理表明イベント (頷き: 頭部の上下の小さな運動 (Pitch)), (d) 回答イベント (頭部の上下の大きな運動 (Pitch)) のイベントが検出可能なことが分かった。図 7 に, 4 問目を出题した際に起こった聞き返し動作 (195 秒前後) の様子を示す。この聞き返し要求時の頭部運動は, 滑らかな曲線を表しており, 二次遅れ系標準形で表現されるマス・ダンパー・スプリングモデルの質量, 粘性摩擦, バネ係数による表現に適合し得ると思われる。つまり, 聞き返し要求をステップ入力として, モデルの質量とバネ係数を変化させることにより曲線のなだらかな動きが表現できる。本実験データは, 視線運動 (頭部運動を含む) 動作生成モデルの基礎的データとなると考える。

5. 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に向けて, 計算課題時の頭部運動データの予備実験と検討を行った。現在我々は, センサや画像処理によって視線方向 (頭部運動を含む) を検出することを検討している。この視線運動検出方法を用いて, 出題者と回答者の両者を撮影・測定対象とすることにより, 対話時における視線運動及び頭部運動の分析につながる重要なデータの抽出が可能と思われる。

今後は, 視線・頭部運動を制御モデルに対応させるために対話時に起こる頭部・視線運動の検出, 発生する各種イベントのモデルへの対応付けを行い, モデルの詳細化および, センサ及び画像処理を応用した頭部・視線運動検出アルゴリズムの研究, 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) および, 科研費特定研究「情報福祉の基礎」(視覚障害者の聴覚認知の解明と音声対話への利用) の支援を受けた。

参考文献

- [Argyle 65] Argyle, M., and Dean, J.: Eye contact, distance and affiliation, *Sociometry*, 28, pp. 289–304, (1965).
- [Argyle 88] Argyle, M.: *Bodily Communication*, 2nd ed. Methuen Co., (1988).
- [荻阪 93] 荻阪良二, 中溝幸夫, 古賀一男: 眼球運動の実験心理学, 名古屋大学出版会, (1993).
- [大坊 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).
- [Duncan 77] Duncan, S.D., Jr., and Fiske, D.W.: *Face-to-face interaction: Research, Methods, and Theory*, Hillsdale, N.J: Lawrence Erlbaum, (1977).
- [福井 84] 福井康之: まなざしの心理学—視線と人間関係—, 創元社, (1984).
- [池田 75] 池田光男: 視覚の心理物理学, 森北出版, (1975).
- [池田 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).
- [Kendon 67] Kendon, A.: Some Functions of Gaze-direction in Social Interaction, *Aca Psychologica*, Vol. 26, pp. 22–63, (1967).
- [Lord 74] Lord, C., Haith, M.M: The perception of eye contact, *Perception & Psychophysics*, 16, 413–416, (1974).
- [中沢 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).
- [嵯峨山 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).
- [田多 91] 田多英興, 山田富美雄, 福田恭介: まばたきの心理学, 北大路書房, (1991).