

音声対話ゲームのためのCGキャラクターの反応的注意生成

星野 准一 森 博志

筑波大学大学院システム情報工学研究科

Generating Reactive Attention of CG Character for Voice Communication Games

Junichi Hoshino Hiroshi Mori

University of Tsukuba, Graduate School of Systems and Information Engineering

{jhoshino, hmori}@esys.tsukuba.ac.jp

アブストラクト

近年、ゲームやテーマパークなどで、CGキャラクターと対話することでストーリーが進行する、音声対話ゲームが見られるようになってきた。このようなゲームでは、ユーザのストーリー体験を強化するために、自然で現実味のある対話を実現することが求められている。しかし、従来ゲームでは、ユーザの発話内容を認識する間CGキャラクターがレスポンスを返すことができないため、CGキャラクターの反応タイミングが遅く不自然に感じられるという問題があった。そのため、本論文では、ユーザの発話内容を理解し反応する熟考的反応だけでなく、会話状態に連動した注意動作や、ユーザの非言語情報を知覚し反射的に注意を向けるといった、反応的注意動作を生成する手法を提案する。最後に、提案手法を音声対話型ゲームに適用し、従来の音声認識の遅延によって発生していたCGキャラクターの反応動作の遅延が軽減され、CGキャラクターとの対話感が増すことを示す。

Abstract

Recently, many interactive systems that talk with CG characters using voice recognition are proposed. However, in the traditional interactive systems, there is a problem that the interval of conversation is unnatural since processing of voice recognition is slow. In this paper, therefore, we propose a technique for generating reactive attention based on user's nonverbal information and a conversation status. In order to realize our technique, we first get verbal data and 3 nonverbal data of a user. And what CG characters pay attention to is decided on those parameters. Then, attention of CG characters is generated by attention angle shearing mechanism. An example of a conversation with CG character using our technique demonstrates the effectiveness of adding attention of CG characters to interactive systems.

1. はじめに

ゲームやテーマパークでみられる、CGキャラクターとユーザが音声対話を行うことでストーリーが進行する音声対話ゲームでは、ユーザのストーリー体験を強化するために、自然で現実味のある対話を実現することが求められている。しかし、従来コンテンツでは、ユーザの発話を認識する間、CGキャラクターがレスポンスを返すことができないため、CGキャラクターの反応タイミングが遅く、不自然に感じられるという問題があった。

実際の人間の行動には、熟考的プロセスと反射的プロセスの

2つのプロセスがあることが指摘され、これらのプロセスによって円滑なコミュニケーションが成り立つとされている[1][2]。熟考のプロセスは、外部からの入力により感情が変化し、最も強い感情に関する入力や目標へ選択的に注意が向けられ、行動が決定される処理である。また、反射的プロセスは、熟考のプロセスでは対応できないような刺激に対して、無意識に反応する処理である。従来研究では、熟考のプロセスに着目し、ユーザの発話内容に対する反応を生成しているものが多いと見られる。しかし、反射的プロセスが十分考慮されていなかったため、音声認識処理によるCGキャラクターの動作遅延が発生

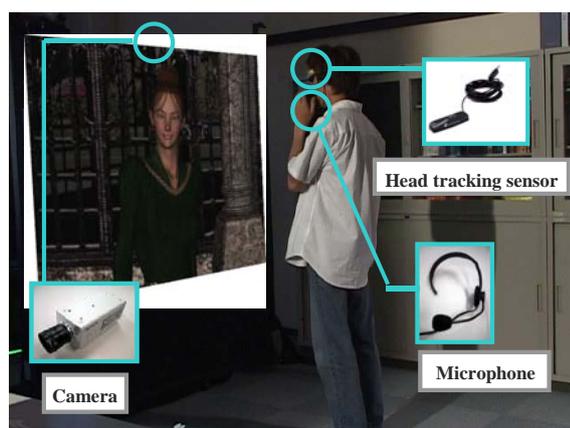


図1. 音声対話型ゲーム

していたと考えられる。

そのため、本論文では、会話状態に連動した注意動作と、ユーザの非言語情報に対する反射的注意動作といった、反応的注意動作を生成することで、音声認識によるCGキャラクターの動作遅延を軽減する手法を提案する。

CGキャラクターの注意動作をリアルタイムで生成するためには、指差しをしながらユーザに注意を向けるなど、特定のモーションを再生している途中で、注意動作を生成する必要がある。そのため、本手法ではCGキャラクターが特定の対象に志向を向ける動作を注意動作とし、頭部、眼球、胴体の回転を既存モーションに上書きすることにより生成する。まず、任意の姿勢や位置関係から注意動作を生成するために、注意を向ける動作を変数により制御するモデルを提案する。次に、角度センサ、マイク、カメラを利用したマルチモーダル環境(図1)において、ユーザの注視位置、発話音量、動作領域の3つの非言語情報と会話状態により、反応的注意動作を自動的に制御する手法を提案する。最後に、提案手法を音声対話型ゲームに適用し、CGキャラクターの動作遅延が軽減され、CGキャラクターとの対話感が増すことを示す。

以下、1.1では関連研究を述べる。2章では本手法の概要を述べ、3章ではCGキャラクターの注意動作の生成について述べる。4章ではユーザ状態の取得について述べ、5章では会話状態に連動した注意動作の制御について述べる。6章では、ユーザの非言語情報に基づいたCGキャラクターの注意動作の生成について述べる。7章では提案した手法の有効性を検証するために、実験により主観評価を行った結果を示す。

1.1 関連研究

実際の人間同士の対話では、発話による言語情報の他に、視線移動[3]や頭部動作[4]といった反射的動作が非常に重要な役割を果たしているとされ[5]、円滑なコミュニケーションの実現に必要な不可欠な要素であることが指摘されている。

視覚に関する反射的注意動作として、対話者の視線や動きなどの特定方向を示す手がかりが、それを察知した者の注意を反射的にシフトさせる、反射的視覚定位があることが分かっている[6]。聴覚による反射的注意動作に関しては、音の鳴った方向へ頭を向ける反射的注意があり、新生児は75%の確率で音のした方向に頭をむけるなど[7]、人間が無意識的に行っている反射的注意であることが分かる。

近年、人間との自然なコミュニケーションを実現するために、そのような機能をCGキャラクターやロボットに応用する研究がおこなわれるようになってきている。有田ら[8]は、視覚的共同注意機能を実装したロボットとユーザとを対話させた実験結果から、視覚的注意機能を実装したロボットのほうが生き物的に見えることを示唆している。著者らは音声対話ゲームにおいて、ユーザの発話を認識すると同時に、ユーザの視線方向からユーザの興味対象を推測し、CGキャラクターが話題提供や視線移動を行う手法を提案している[9]。Chapmanらの研究では、特定の情報に注目する注意機構を実現し、話題に関連した単語やオブジェクトに注意を向けることによって発話内容が限定され、応答文の生成速度を向上させることができることを示している[10]。

これらの知見から実際の人間に近い注意動作をさせることで、人間らしい動作生成が行われ、より自然な対話が可能であることがわかる。しかし、従来研究はいずれも、注意動作の生成手法に注目しているため、反射的注意の生成と音声認識処理とを併用することで、音声認識によるCGキャラクターの動作遅延の軽減につながるのかどうかという点については考察されていない。

2. 概要

本手法では、音声対話型ゲーム(図1)を用いて、CGキャラクターとユーザが対話を行う。システムには、ヘッドトラッキングセンサ、マイク、カメラが内蔵されており、ユーザの発話認識の他に、ユーザの注視位置、発話音量、動作領域がリアルタイムで検出される。これらのデバイスから得られた非言語情報を参照し、注意モジュール(Attention Module)が反射的な注意動作を生成する(図2)。

また、実際の会話では聴取時や発話時といった会話の状態によって、対話者へ注意を向ける頻度や長さが異なることから[5]、会話の状況に応じた注意動作を注意モジュールが確率的に生成する。本手法では、CGキャラクターの発話の有無により、発話状態、聴取状態、2者間に発話がない待機状態の3つの状態に分類し(図3)、注意モジュールが会話状態を参照する。これにより、ユーザの非言語情報と会話状態とに連動した注意動作を生成することができる。

従来の注意動作の生成では、あらかじめ作成されたモーションデータを再生するだけであったため、多様な位置関係で、任意の対象に注意を向けることが困難であった。また、特定のモーションと並行した注意動作を生成することが望まれている。そのため、注意方向を、注意分担機構(Attention Shearing Mechanism)により頭部、眼球、胴体の回転に分解し、既存モーションに合成することで注意動作を生成する。これにより、特定のモーションと並行した注意動作が可能になる。

転角の割合を分担比と定義し、その算出方法について述べる。

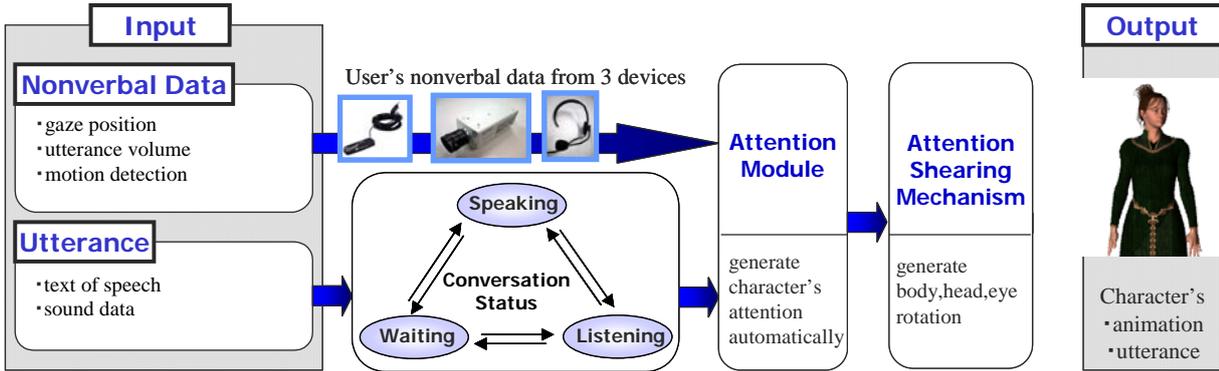


図 2. 反動的注意動作の生成概要

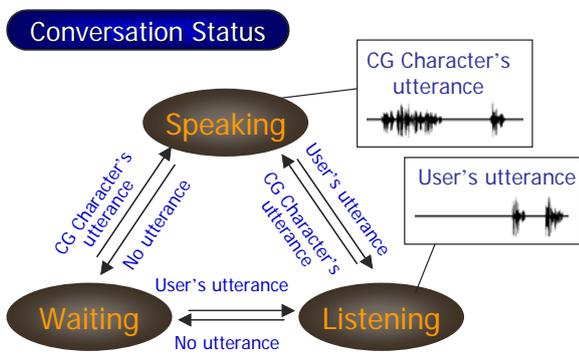


図 3. 会話状態の遷移

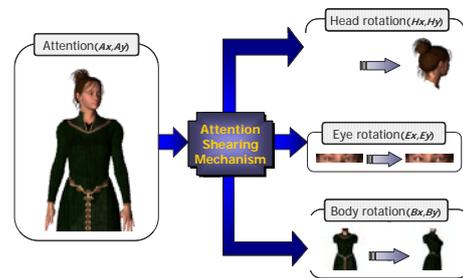


図 4. 注意動作の分担

3. 注意動作の生成

CGキャラクタの注意動作をリアルタイムで生成するためには、指差しをしながらユーザに注意を向けるなど、特定のモーションを再生している途中で注意動作を生成する必要がある。そこで、著者が既に提案した、頭部・眼球的動的分担機構[11]を拡張し、眼球・頭部・胴体の動的分担機構を用いた注意動作の生成を行う(図4)。

頭部・眼球的動的分担機構[11]とは、CGキャラクタの視線方向を頭部と眼球的回転角の和とし、視線方向によって頭部・眼球的回転角の割合を動的に変更するモデルである。本手法では、従来モデルよりも広範囲に注意を向けるため、注意方向を眼球、頭部、胴体の回転角の和として定義する。また、従来モデルでは視線方向のみによって各部位の回転角の割合を決定しているが、実際の会話では、各部位の回転の割合は様々な要因によって決定されるものと考えられる。そこで、本研究では、より自然な対話中の動作を表現するために、注意方向だけでなく、CGキャラクタがユーザから受ける刺激も考慮して、各部位の回転角の割合を動的に変化させる。なお本論文では、正面を向いた状態の角度を 0° 、左を向く角度を正、右を向く角度を負の値とし注意方向を表現する。

3.1 注意方向による分担比

本節では、注意方向に対する眼球・頭部・胴体それぞれの回

ただし、 A_x, A_y をそれぞれ注意方向の水平角成分、垂直角成分とし、同様に E_x, E_y を眼球、 H_x, H_y を頭部、 B_x, B_y を胴体のもとする。また、眼球的回転角の水平方向、垂直方向の分担比をそれぞれ D_{Ex}, D_{Ey} 、同様に D_{Hx}, D_{Hy} 、を頭部、 D_{Bx}, D_{By} を胴体の分担比とする。

$$\begin{bmatrix} A_x \\ A_y \end{bmatrix} = \begin{bmatrix} E_x + H_x + B_x \\ E_y + H_y + B_y \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} E_x & E_y \\ H_x & H_y \\ B_x & B_y \end{bmatrix} = \begin{bmatrix} D_{Ex} & D_{Ey} \\ D_{Hx} & D_{Hy} \\ D_{Bx} & D_{By} \end{bmatrix} \begin{bmatrix} A_x & 0 \\ 0 & A_y \end{bmatrix} \quad (2)$$

従来研究により、視線移動は、 $30^\circ \leq |V_x|, 20^\circ \leq |V_y|$ の場合は頭部の回転のみで目標を捉えることが分かっている[11]。ただし、 V_x, V_y はそれぞれ視線の向きの水平角成分、垂直角成分を表す。本論文で扱っている注意動作時には、視線移動時よりも広範囲に志向をむけるため、胴体の回転を伴った回転動作が行われ、一定の角度を超えると注意を向けやすい場所へ立ち位置を移動することも行われる。そこで、注意動作における胴体と頭部の関係、立ち位置の関係を調査するために、注意動作における各部位の回転角のデータを実際に計測し考察した。その結果、

- 1) 左右ともに注意方向が水平角 130° を超える視線移動時には、より見やすい位置へ立ち位置を変更する。
- 2) 左右とも二注意方向が水平角 50° を超える視線移動時には胴体の回転を伴い、水平角 130° では頭部と胴体の回転

角はほぼ等しい。

- 3) 垂直方向の視線移動には胴体による回転はほぼ見られない。
- 4) 立ち位置の変更は右側(負の角度)を向く場合は右回転、左側を向く場合は左回転となる。

ということが明らかになった。以上の考察の結果から、水平方向成分については、 $0^\circ \leq |A_x| < 15^\circ$ では眼球のみの回転、 $15^\circ \leq |A_x| < 30^\circ$ では、眼球と頭部の回転、 $30^\circ \leq |A_x| < 50^\circ$ では頭部のみの回転、 $50^\circ \leq |A_x| \leq 130^\circ$ では頭部と胴体の回転により水平方向の注意動作を行う。 $130^\circ < |A_x|$ の場合は、より見やすい位置へ立ち位置を変更するものとする。

$$D_{Ex} = \begin{cases} \alpha & (0 \leq |A_x| < 15) \\ \alpha \left(\frac{30 - |A_x|}{15} \right) & (15 \leq |A_x| \leq 30) \\ 0 & (30 < |A_x|) \end{cases} \quad (3)$$

$$D_{Hx} = \begin{cases} \beta(1 - D_{Ex}) & (0 \leq |A_x| < 50) \\ \beta \left(1.0 - \frac{0.5 \cdot (|A_x| - 50)}{80} \right) & (50 \leq |A_x| \leq 130) \\ 0 & (130 < |A_x|) \end{cases} \quad (4)$$

$$D_{Bx} = 1 - D_{Ex} - D_{Hx} \quad (0 \leq |A_x| < 130) \quad (5)$$

垂直角成分については、 $0^\circ \leq |A_y| < 10^\circ$ では眼球のみの回転、 $10^\circ \leq |A_y| \leq 20^\circ$ では眼球と頭部の回転、 $20^\circ < |A_y|$ では頭部のみの回転で垂直方向の注意動作を行う。また、垂直各方向の注意移動に関しては胴体による回転はほとんど見られなかったため、垂直捕獲方向における胴体の回転は考慮しないものとする。

$$D_{Ey} = \begin{cases} \alpha & (0 \leq |A_y| < 10) \\ \alpha \left(\frac{20 - |A_y|}{10} \right) & (10 \leq |A_y| \leq 20) \\ 0 & (20 < |A_y|) \end{cases} \quad (6)$$

$$D_{Hy} = \beta(1 - D_{Ey}) \quad (0 \leq |A_y|) \quad (7)$$

$$D_{By} = 1 - D_{Ey} - D_{Hy} \quad (0 \leq |A_y|) \quad (8)$$

ただし、分担比は個人や年齢によって差が大きい[12][13]、本稿では重み α ($0 \leq \alpha \leq 1$)、 β ($0 \leq \beta \leq 1$) により調整可能にする。 $\alpha = \beta = 1$ のときの分担比による眼球・頭部・胴体の水平角成分を図5に、垂直角成分を図6に示す。

3.2 重要度による分担比の制御

実際の対話では対話者の注意が大きく変化した場合、注意がどこに向いているかは眼球方向のみならず、頭部や身体姿勢といった多様な手掛かりから推定することができる。またWollastonらの研究によれば頭部の向きにより他者の注意がどこに向いているかを判断することが難しくなることが指摘されている[14]。

そのためキャラクターがどこに注意を払っているのかをユーザに正確に伝えるためには、キャラクターの注意の度合いによって頭部の動きを強調させる必要がある。

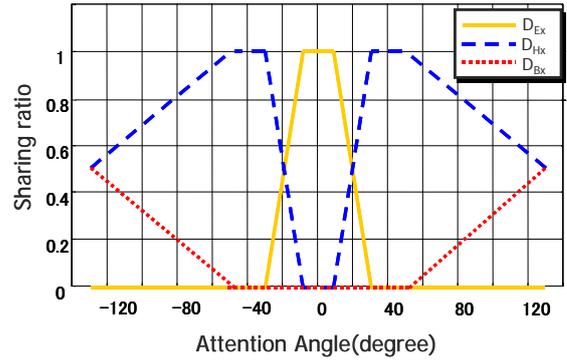


図5. 分担比による眼球・頭部・胴体の水平角成分

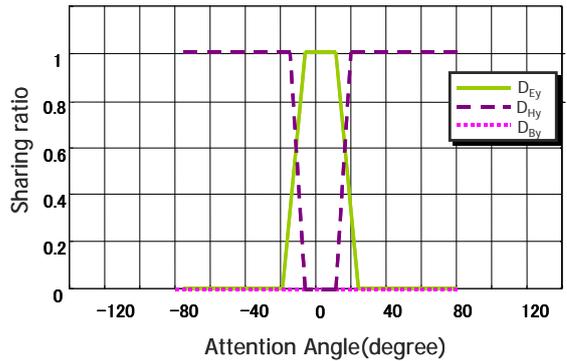


図6. 分担比による眼球・頭部・胴体の垂直角成分

そこで本手法では4章で述べる注視点の重要度により3.1で述べた分担比の重み α を変化させることで、キャラクターの注意度合いを表現する。重要度は大中小の3段階とし、それぞれ α の値を、0.5, 0.75, 1.0とした。これにより注視点の重要度が高いほど視線移動における頭部の回転の割合が多くなり、ユーザにキャラクターの注意の所在を正確に認識させることが可能となる。

3.3 頭部・眼球・胴体の遷移モデル

本手法では、注意移動における眼球・頭部・胴体の回転動作をモデル化するために、典型的な視線移動[15]と注意移動[16]を参考にし、注意移動における頭部・眼球・胴体の遷移モデルを提案する。図7は視線を視線方向 $A(t)$ と胴体、頭部、眼球それぞれの回転 $B(t)$ 、 $H(t)$ 、 $E(t)$ の時間変化を示している。ただし、図7中の E_{max} を眼球の限界角度、 H_{Th} 、 B_{Ta} をそれぞれ分担比によって算出された頭部、胴体の角度とする。このモデルは以下の4つのステップからなる。

- 1) 眼球の角速度は頭部・胴体の角速度よりも非常に大きいいため眼球が先に動きだし、サックード後に眼球の角速度が眼球の限界角に到達する。
- 2) 頭部が回転を始める。頭部は分担比によって計算された角度になるまで回転し続け $B(t)$ 、 $H(t)$ 、 $E(t)$ の和が視線方向 V となるまで眼球は最大角を保ち続ける。
- 3) 頭部の回転開始後に胴体の回転が発生し、分担比によって計算された角度になるまで回転し続ける。
- 4) 眼球がしばらく最大角を保ち続けた後に、頭部、胴体の

回転に伴って、眼球が逆方向に回転し始める。

ステップ2の頭部の回転は角速度をシグモイド状に近似できることが知られている。本稿ではステップ3における胴体の回転の角速度変化もシグモイド状に近似できるものとした。

4. ユーザの非言語情報の検出

本論文では、CGキャラクターの視覚的注意動作と聴覚的注意動作を生成するために、1)ユーザのスクリーン上の注視位置、2)単位時間内に大きく動いた動作箇所、3)発話の音量という3つの非言語情報を、それぞれ頭頂部に取り付けられたヘッドトラッキングセンサ、カメラ、マイクロフォンを用いてリアルタイムに検出する。発話音量についてはマイクから簡易に取得できるため本章では省略する。

4.1 ユーザの注視位置の推定

対話中の視線行動は非常に重要な役割を持つことが指摘され[3]、実際の人間は対話者の注意方向へ不意に注目するといった現象がある。そのような反射的注意を実現するために、ユーザの注視位置を推定する。

本論文で扱う大型スクリーン上でのインタラクティブ環境では、頭部の動きが主となる視線移動が多くみられた。そこで、本論文ではユーザのスクリーン上の視点を頭部の角度であると近似し、比較的安価で分解能の高いヘッドトラッキングセンサを用いて、頭部の角度から注視位置を推定する(図8)。

注視位置は、検出された頭部回転角から、以下の式により算出される。

$$\begin{bmatrix} x_{gaze} & y_{gaze} \end{bmatrix} = \begin{bmatrix} (W_w / 2 + x_0) + x_1 & (H_w / 2 - y_0) - y_1 \end{bmatrix} \quad (9)$$

$$\begin{bmatrix} x_1 & y_1 \end{bmatrix} = \begin{bmatrix} \frac{W_w / 2 \times \theta_x}{\theta_{x_max}} & \frac{H_w / 2 \times \theta_y}{\theta_{y_max}} \end{bmatrix} \quad (10)$$

$$\begin{bmatrix} \theta_{x_max} & \theta_{y_max} \end{bmatrix} = \begin{bmatrix} \tan^{-1}(\frac{L - H_e}{D}) & \tan^{-1}(\frac{W_s / 2 - P}{D}) \end{bmatrix} \quad (11)$$

$$\begin{bmatrix} x_0 & y_0 \end{bmatrix} = \begin{bmatrix} \frac{P \times W_w}{W_s} & \frac{H_w \times \{H_e - (L - H_s)\}}{H_s} \end{bmatrix} \quad (12)$$

ただし、 W_s , H_s を実世界でのスクリーンの横幅、縦幅、 W_w , H_w をそれぞれアプリケーションウィンドウの横幅、縦幅、 L を地面から測った画面の高さ、 D をユーザから画面までの距離、 H_e を地面から測ったユーザの目の高さ、 P を画面の中心から見たユーザの左右の立ち位置、 x_0 , y_0 を初期の注視位置のx座標、y座標、 θ_x , θ_y を頭部のx軸周りの回転角、y軸周りの回転角、 θ_{x_max} , θ_{y_max} を頭部がスクリーンの上端を向いた時の角度、スクリーンの右端を向いた時の角度、 x_1 , y_1 を頭部の角度によって決まるx座標の増減、y座標の増減、 x_{gaze} , y_{gaze} を最終的に求まる注視位置のx座標、y座標とする。ユーザの視線の停留時間が長いほど、注視点の重要度が高くなるものとし、注意移動を行う際の各部位の分担比に影響する。

4.2 ユーザの動作箇所の検出

実際の人間の対話では、対話者が頭に触れるしぐさや、急に

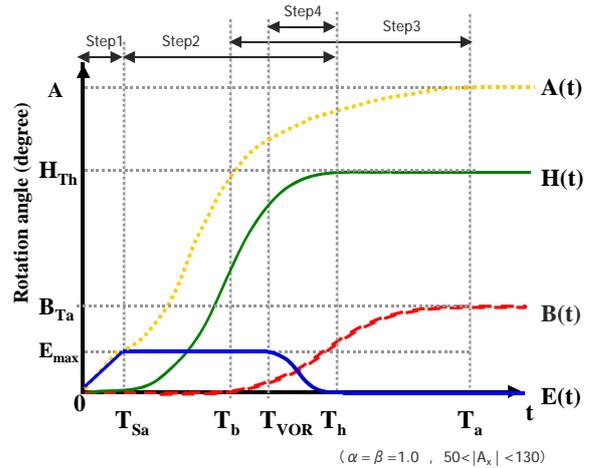


図7. 注意動作における頭部・眼球・胴体の時間特性

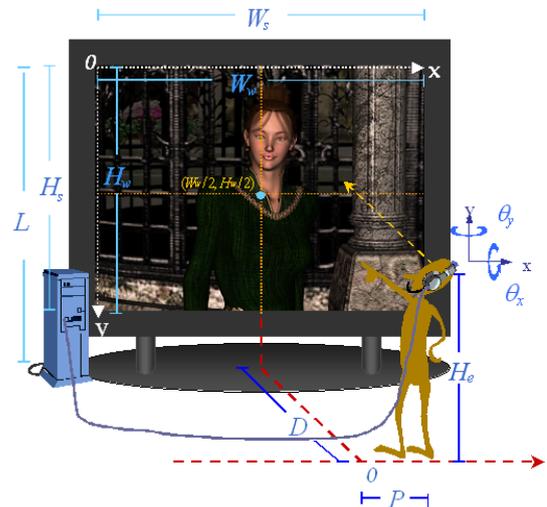


図8. ユーザの注視位置の推定

動かした手の方向を反射的に見ることがある。これらは、単位時間内に大きく動いた動作箇所の方に、反射的に注意を向ける動作である。本手法では画像中の肌色領域を認識し、単位時間あたりの動作量が閾値を超えた箇所を検出する。

動作箇所の検出は以下の手順により求める。

- 1) フレーム間差分処理により動作領域を抽出する。得られたフレーム画像 $f(m, n)$, $f_{i+1}(m, n)$, ... をグレースケールに変換し差分を求めた。

$$S_i(m, n) = |f_{i+1}(m, n) - f_i(m, n)| \quad (13)$$

- 2) カメラから得られた画像中の肌色領域を抽出する。肌色領域は色相がほぼ一定であるため、以下の式により色相を求め閾値処理により肌色領域を抽出する[17]。

$$H = \cos^{-1} \left\{ \frac{(G-B) + (G-R)}{\sqrt{(G-B)^2 + (G-R)(B-R)}} \right\} \quad (14)$$

- 3) 2)で求めた肌色領域の重心を求め、大まかな動作箇所の特定をする(図9)。

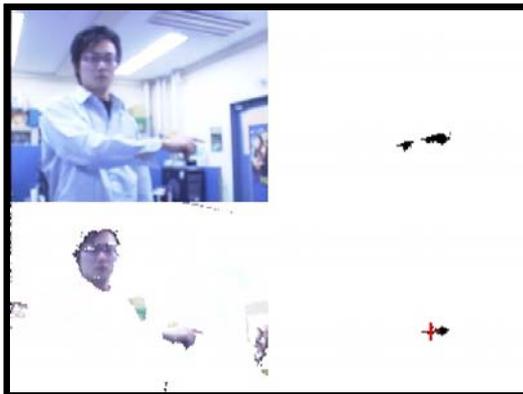


図9. ユーザの動作箇所を推定

図9中の左上の画像は元画像，右上がフレーム間差分によって得られた動作領域，左下は色相による肌色の検出領域，右下が検出されたユーザの動作箇所を表している。

単位時間あたりの動作量が大きい動作箇所ほど重要度が高いものとし，動作量が閾値を超えた箇所に対してCGキャラクタは注目する。

5. 会話状態に連動した注意動作の生成

5.1 CG キャラクタの注意状態

本手法では，CGキャラクタの注意がユーザに向いている状態 S_u と，ユーザから注意をそらしている状態 S_o の2状態を考え，注意率(R)と注意持続時間 (L [ms]) の2変数を用いて注意状態を制御する。 R は対話時間中にユーザに注意を向けていた時間の割合， L は1回の注意が持続する平均の時間とする。また，状態 S_u での注意の対象となる座標を C_u ，状態 S_o での注意の対象の座標を C_o とする。ただし，抽出した肌色領域中の最も上の領域を，ユーザの顔の領域であるとみなし， C_u の座標とする。座標 C_o は会話状態や非言語情報によって決定される。

変数 R ， L に従った眼球移動を生成するために2状態 (S_u ， S_o) の確率遷移モデルを考える。 S_o から S_u に遷移する確率 P_{ou} を， S_u から S_o に遷移する確率 P_{uo} を， S_u から S_u に遷移する確率 P_{uu} を， S_o から S_o に遷移する確率を P_{oo} と定義すると各状態遷移率は以下のように定義される。

$$P_{ou} = P_{uu} = R \quad (15)$$

$$P_{oo} = P_{uo} = 1 - R \quad (16)$$

状態が S_u に遷移してから次に遷移するまでの平均時間を L に一致させるために，状態遷移が発生する周期 t [ms]を以下の式によって定義する。

$$t = (1 - R) \times L \quad (17)$$

ただし，本稿では $R=0.5$ ， $L=1000$ [ms]を注意率，注意持続時間の標準値とする。また， $R=1.0$ が常にユーザに注意を向けることを表し， $R=0.0$ が常にその他の方向に注意が向くことを表す。

表1. 会話状態による注意状態パラメータの動的設定

会話状態	注意率:R	注意持続時間:L[ms]
待機状態	0.5	1000
発話状態	0.5	1000
発話状態の開始部	0.75	1000
発話状態の終了部	0.75	1000
聴取状態	0.75	1500

5.2 会話状態に連動した注意移動

社会心理学の分野では，会話状態により表出される動作が変化するとされ，Argyleの実験によると，

- (a) 聴取状態の方が相手を凝視する頻度が高い。
- (b) 聴取状態の方が相手を凝視する時間が長い。
- (c) 発話者は話の区切りや話の終了時に凝視することが多い。

ということが明らかになっている[18]。しかし，5.1で述べた各変数の標準値は，会話状態を考慮にいれていないため，(a)~(c)の条件を十分に満たすことができない。そのため，本手法ではユーザとCGキャラクタの発話の有無により得られた会話状態を用いて，各変数を動的に変更し，会話に連動したCGキャラクタの注意動作の制御を行う。

本手法では，条件(a)を反映させるためにCGキャラクタが聴取状態である場合に注意率 R を増加させ，条件(b)を反映させるためにCGキャラクタが聴取状態である場合に注意持続時間 L を増加させる。さらに条件(c)を反映させるために，CGキャラクタが発話状態である場合に発話開始部と発話終了部における注意率を増加させる (表1)。

これにより，会話状態を考慮したCGキャラクタの注意動作の制御が可能となり，無意識的に視線を変化させる身体操作や，会話の始めに視線を交わす規制的動作を表現することができる。

6. 反射的な注意動作

人間は対話している相手をモニタする機能があることが指摘され[3]，対話者の視線の方向に不意に注意を向けることや，声を感じて相手の方を向くことなどは，反射的な注意動作である。本章では，ユーザの非言語情報をもとに，CGキャラクタの反射的な注意動作を生成する手法について述べる。

6.1 ユーザの注視位置による注意

対話中の視線行動は非常に重要な役割を持つことが指摘され[3]，人間は対話者の視線の方向へ不意に注目したり，自分に対して程度の視線が注がれると，反射的に目をそらすことがある。

そこで，ユーザの注視位置によって以下の2つの反射的な注意動作を生成する。

- 1) 通常より長時間目を合わせている場合，ユーザから注意をそらす
- 2) ユーザの注意方向を気にする (共同注視)

本手法ではCGキャラクタの注意の移動を自動的に生成しているため，その際に用いる各変数を，ユーザの注視位置をもとに動的に変更し，ユーザの注意を意識した注意動作を生成する。

- 1)の視線を外す動作は，状態 S_u かつ，ユーザの注視位置がCG

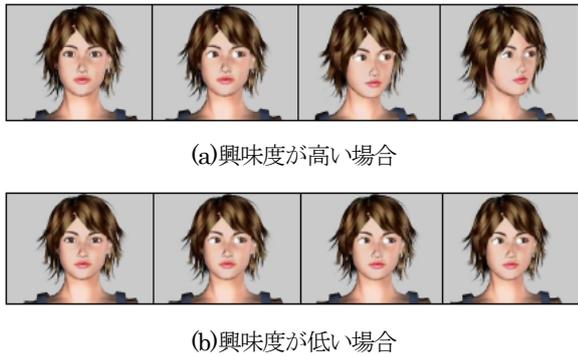
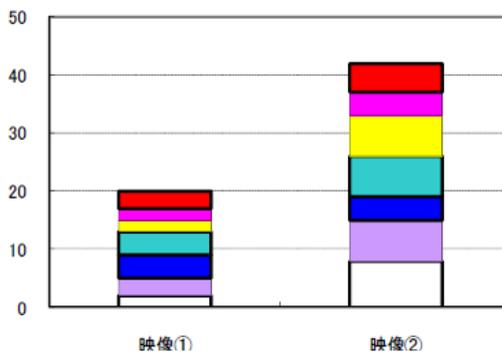


図 10. 興味度による頭部動作の違い



- 会話のやりとりがスムーズに感じる
- 実際の人間同士で話しているように感じられる
- 会話中の動作が自然に見える
- 眼球や頭部や胴体の動きが自然である
- 会話の間が自然である
- 動作や行動が実際の人間らしい
- CGキャラクタが相手の様子を察知しているように感じる

図 11. アンケート結果

キャラクタの顔付近に1秒以上停留した場合、ユーザから視線を回避する確率を高くすることで生成する。これらの変数値は、対話における二者間での相互の顔面への凝視する相互注視は約1秒であるという結果から決定した[18].

2)の共同注視は、状態 S_0 での注意対象の座標 C_0 をユーザの注視対象の座標に変更し、ユーザの視線方向を気にするような共同注視動作を生成する。ただし、常にユーザの注視対象を状態 S_0 での注意方向とした場合、ユーザの注視位置に注意を向ける確率が高くなるため、ユーザの注視位置が特定の範囲に2秒以上停留した場合に座標 C_0 を変更する。

6.2 動作箇所への注意

人間は実際の対話中に、話し相手が頭に触れるしぐさや、急に動かした手の方向を反射的に見ることがある。これらはいずれも、単位時間内に大きく動いた箇所である動作箇所の方向に、反射的に注意を向ける動作である。そこで、ユーザ領域が大きく動いた場合、CGキャラクタがその箇所に注目するという行

動を生成する。

ユーザの動作箇所は4.2で述べた手法を用いて検出し、状態 S_0 での注意対象の座標 C_0 を、検出された動作箇所の座標とし、 $R=0.1$, $L=1000$ [ms]とすることで、ユーザの動作箇所へ注意動作を生成する。また、このような動作はユーザの急激な動作に対して、反射的に反応する注意動作であるため、3.2で述べた注意を向けるまでの所要時間であるパラメータ T_0 の値を通常よりも小さくする。

6.3 ユーザの発話による注意

実際の対話において、人は対話者に注意が向いてなくても、対話者が発話すれば自然と相手に注意を向ける。また、発話音量が大きい場合は、より素早い注意動作がおこなわれる。そこで、ユーザの発話音量によってCGキャラクタの反射的注意動作を生成する。一般的に音量の取得は音声認識の処理速度と比べて非常に高速であるため、反応的注意動作を生成する要因として有効である。ユーザの発話が入力された瞬間に、その音量が閾値以上ならば、その直後の注意率、注意持続時間をそれぞれ $R=1.0$, $L=1500$ [ms]とする。ユーザの適切な音量での発話が続いている期間は、会話状態が聴取状態であるので、発話の音量による注意の跡は聴取状態での注意状態の遷移に従う。また、声の大小に反応する反応的注意動作は、入力された音量のレベルによって注意を向けるまでの所要時間であるパラメータの値 T_0 を上下させる。

7. 評価実験

7.1 反応的注意動作の検証

本手法を用いることで、ユーザの発話に対するCGキャラクタの反応の遅さが解消され、対話への没入感が高まるかどうかの検証を行った。音声対話ゲームを使い、提案手法をCGキャラクタに適用しない場合(映像①)と、適用した場合(映像②)とを実際に体験し、アンケートにより評価を行った。実験では、「中世ファンタジーの世界の武器屋で、ユーザが女性店員役のCGキャラクタに槍の値段を尋ねる」というシーンを設定し、

- ユーザ：「すみません」
- CGキャラクタ：「いらっしやいませ」
- ユーザ：「その槍はいくらですか？」
- CGキャラクタ：「この槍は15,000Gになります」

という短い会話を行った。映像①はあらかじめ用意した動作を使用し、動作同士を補間して再生するだけのアニメーションである。映像②は映像①に提案手法を適用し、発話に応じた動作だけでなく、反応的注意動作を同時に生成した映像である。今回の実装ではより自然な対話を実現するために、3.2節で述べたように、重み α を注視点の重要度として扱い、重要度の変化に応じて、頭部動作の分担比を変化させる(図10)。これにより、状況に応じた適切な頭部動作を生成することが可能となる。さらに6.3節で述べたように、入力された音量のレベルに応じて、

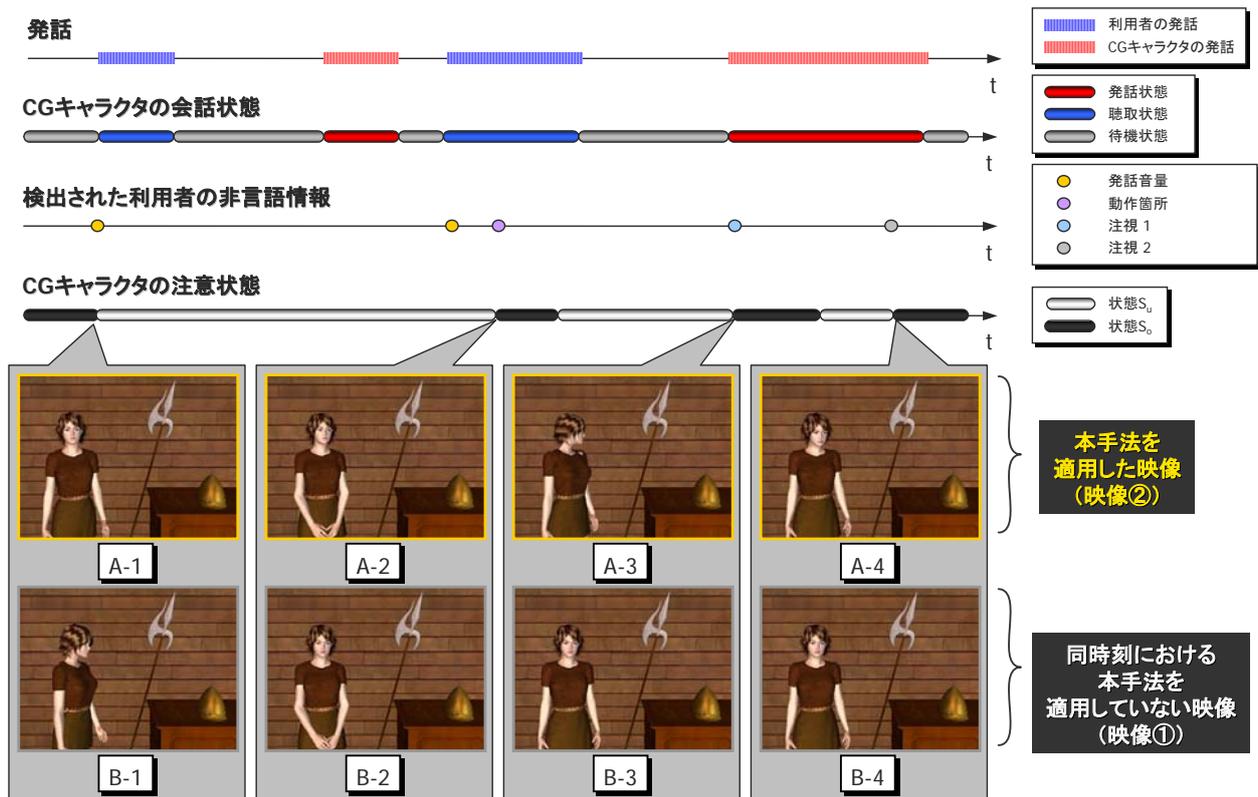


図 12. 実験結果

注意を向けるまでの所要時間を変化させた。

実際に被験者7人に2つのアニメーションを提示し、一対比較を行った。評価方法は、2種類のアニメーションを比較するために評価項目をマークする形式であり、評価項目ごとに点数を算出し、すべての点数を合計したものを各々の映像の得点とした(図11)。点数は、「非常に①がよい」だった場合、映像①にのみ2点を加え、「どちらかという②がよい」だった場合は、映像②にのみ1点を加え、「どちらでもない」場合は点数を加えないという方法で計算した。

図11に示す結果より、反応的注意動作の付加した本手法の方がCGキャラクタとの対話感が全体的に向上していることが分かる。「会話の間が自然である」の項目については、会話音声と音声に伴う動作については同一のデータを用いているため印象が変化していないことに加えて、反応的注意動作が音声に伴う会話の間について及ぼす影響が少ないことがわかる。

図12下部に、実験で用いた映像②のタイムラインと静止画像、同時刻における映像①の静止画を示す。図12上部はユーザとCGキャラクタの発話、CGキャラクタの会話状態、検出されたユーザの非言語情報、CGキャラクタの注意状態のタイムラインを表している。図12中の注視1はユーザの注視位置が単位時間内で特定の範囲に閾値以上時間底流することを、注視2はユーザの注視位置が1秒以上CGキャラクタの顔付近に手移駐することを表す。

図12-A-1はユーザが適切な音量で「すみません」と発話したため、CGキャラクタが反応し、ユーザに注意を向けている様

子である。図12-A-2ではユーザの動作箇所を検出したことで、CGキャラクタの動作箇所への注意が生成されている。図12-A-3はユーザの注視位置が壁際の槍に閾値以上の時間底流したため、CGキャラクタの共同注視が生成され、図12-A-4はユーザの注視位置が1秒以上CGキャラクタの顔付近に底流したため、ユーザから注意をそらす動作が生成されている。図12-B-1～4では反応的注意動作が生成されず、音声認識処理中は、ユーザに対して反応を全く返していない様子が確認できる。

8. おわりに

本稿では、従来の音声認識によって発生していたCGキャラクタの動作遅延をなくし、より自然で現実感のある対話を実現するために、会話状態とユーザの非言語情報とを反映した、CGキャラクタの反応的注意動作の生成について述べた。

その実現のために、検出までに時間がかかる言語情報と並行して、素早く検出できるユーザの注視位置、動作箇所、発話音量の3つの非言語情報と会話状態を利用した。これらをCGキャラクタに反映させて注意状態を決定することにより、会話状態に連動した注意動作と、反射的な注意動作を自動的に生成した。また、検出される情報を柔軟に反映させるために、注意動作の生成にはパラメータ制御による眼球・頭部・胴体の動的分担機構を提案した。

評価実験では、CGキャラクタが反応的注意動作を行うことで、あたかもCGキャラクタがユーザの発話を即座に知覚して反応しているように見せることが可能となり、対話感の向上に

有効であることが確認できた。

今後の課題として、実際の人間の対話シーンと本手法で生成されるアニメーションを比較し、対話への没入感を向上させる要素を明確にし、実装することが必要である。また、注意動作に伴う身体部位へ振動の影響を考慮するなど、より高精細な注意動作を生成するモデルを提案することが必要である。

謝辞 本論文を作成するにあたり、協力して頂きました益子宗さん、長谷将生さんに御礼を申し上げたく、謝辞に代えさせていただきます。

参考文献

- [1] 戸田正直, 感情, 東京大学出版, 1992.
- [2] A.R.Damasio, *Descartes' Error*, New York, 1994.
- [3] Kendon, A., Some functions of gaze-direction in social interaction, *Acta Psychologica*, Vol.26, pp.22-63, 1967.
- [4] H. Ogawa, T. Watanabe, *InterRobot: speech-driven embodied interaction robot*, *Advanced Robotics*, Vol.15, No.3, pp.371-377, 2001.
- [5] Mehrabian, A., *Communication without words*, *Psychological Today*, 2, pp.53-55, 1968.
- [6] reflexive visual orienting :Driver et al., 1999; Friesen & Kingtone, 1968; Hood et al, 1998; Langton & Bruce, 1999
- [7] Muir, D.W., Clifton, R.K., &Clarkson, M.G., 1989, The development of a human auditory localization response; A u-shaped function, *Can J Psychol*, 43, 199-216.
- [8] 有田亜希子, 開一夫, 小松孝徳, 松田剛, 松本吉央, 人間と機械の共同注意に関する研究, 情報処理学会第126回知能と複雑研究会(SIG-CIS), pp.204-205, 2001.
- [9] 益子宗, 星野准一, CGキャラクタのための視線移動モデルの構築, NICOGRAPH Spring Festival in TAF, 2007.
- [10] Chapman, D., *Vision, Instruction, and Action*, MIT press, 1991.
- [11] 益子宗, 星野准一, 仮想俳優のための頭部・眼球動作の生成, 信学論, Vol.J88-D-II, No.3, pp.585-595, 2005.
- [12] Tabuchi, A., Fukushima, M. and Kajikawa, I., *Eye-Head coordination in Children*, *Ganki*, Vol.34, pp.1212-1215, 1983.
- [13] Yamada, M., *The analysis of head and eye-coordination according to age*, *NHK STRL R&D No.44*, pp.29-37.
- [14] Wollaston, W.H., *On the apparent direction of eye in a portrait*, *Philosophical Transactions of the Royal Society of London*, B114, pp.247-256, 1824.
- [15] Tweed, D., Fetter, M., Sievering, D., Misslisch, H. &Koenig, E., *Rotational kinematics of the human vestibule ocular reflex II :Velocity steps*, *J.Neurophysiol.* 72, pp.2480-2489, 1994.
- [16] 飯島貴志, 人体のしくみ, ワークスコーポレーション, 2003.
- [17] 松橋聡, 藤本研司, 中村納, 南敏, 顔領域中衆ルに有効な修正HSV表色系の提案, *テレビジョン学会誌*, Vol.49, No.6, pp.787-797, 1995
- [18] Argyle, M. and Dean, J., *Eyecontact, distance and affiliation*,

Sociometry, 28, 289-304, 1965.

星野 准一



平成元年, 早稲田大学工学部電子通信学科卒. 同年, セコムIS研究所入所. 平成3年から平成5年 MITメディアラボ客員研究員. 平成11年から新潟大学自然科学研究科助手. 平成12年から筑波大学機能工学系講師. 平成12年~15年科学技術振興事業団さきがけ研究21「情報と知」領域 研究員を兼務. 平成14,15 ACM Advances in Computer Entertainment, Program Chair, 平成15~16 日本バーチャルリアリティ学会エンタテインメントVR研究委員会委員長, 平成17~18 情報処理学会エンタテインメントコンピューティング研究会主査. 工学博士. エンタテインメントコンピューティング, ゲームテクノロジーなどの研究に従事. IEEE, ACM,情報処理学会, 各会員.

森 博志



平成14年, 筑波大学卒, 平成19年, 同大学大学院終了. 平成20年から同大学大学院システム情報工学研究科研究員. 工学博士. エンタテインメントコンピューティングなどの研究に従事.