

音声駆動型身体引き込みキャラクターの振り向き動作による 子どもの注意獲得・維持システム

A Children's Attention Attracting/Continuing System

by Turning Motions of the Voice-Driven Embodied Entrainment Character

久富 彩音[†] 石井 裕^{††} 渡辺 富夫^{††}

Ayane Hisatomi[†] Yutaka Ishii^{††} Tomio Watanabe^{††}

[†]岡山県立大学大学院 情報系工学研究科 ^{††}岡山県立大学 情報工学部

1 はじめに

ロボットの振り向き動作に関する研究 [1] では、ロボットがユーザの存在を検知し、ユーザに対して視線を向けることで、ユーザの話聞く意思があることを認知させることが知られている。一方、ユーザに対する注視行動を常に行うのではなく、ユーザを注視する時間が対話時間の80%になるように視線を外すことで、ユーザが感じる威圧感や圧迫感を軽減する効果があることが明らかにされている [2]。これらの知見から、ロボットの視線がユーザに及ぼす心理的影響が、二者間のインタラクションにおいて重要であると考えられる。

これまで、著者らは iRT (音声駆動型身体引き込み技術) を CG キャラクタ等に実装することで、ユーザへのコミュニケーション支援を行うシステムを開発し、その有効性を示している [3]。iRT は人の発話音声のリズムに基づくモデルであり、その特性からユーザがキャラクターに呼びかけるような短い発話では応答反応を示さない場合がある。この場合、キャラクターが音声で反応することをユーザは認知できず、ユーザからのキャラクターへの注意を獲得できない可能性がある。とくに、小学生以下の子どものユーザには、挨拶や単語といった比較的短い発話を行う傾向がみられ、従来のシステムを使用した際にシステム内のキャラクターが応答反応を返さないことが想定される。また、子どものユーザに対し、システム使用時の発話の継続を期待するためには、ユーザがキャラクターに向ける注意の維持が課題であると考えられる。

本研究では、小学生以下の子どものユーザの発話促進、維持を目的とし、従来の iRT による聞き手のうなずき動作に加え、ユーザの音声の入力開始時にキャラクターが振り向き動作を行うことで、キャラクターが自身の音声に反応することを継続的に認知させ、ユーザの注意を獲得・維持するシステムを開発した。また、小学3~6年生を対象とした発話実験により、開発したシステムの有効性の検証を行うとともに、システム使用時の視線情報の解析を行った。

2 開発システム

2.1 コンセプト

本研究で開発したシステムのコンセプト図を図1に示す。システム使用開始後、聞き手キャラクターはシステム画面内の別オブジェクトに興味を示すような遊戯行動を行う。従来の対話システムでは、ユーザからの積極的な発話を前提としていたため一貫してキャラクターがユーザに正対していたのに対し、遊戯行動を行うことで、積極的にキャラクターへの発話をする意思がないユーザの注意を獲得することができる。また、聞き手キャラクターはユーザの発話時間に関係なく、音声のON-OFFに反応してユーザの方向へ振り向き動作を行う。短時間の発話に対してもキャラクターが反応を示すことで、ユーザはキャラクターが自分の音声に反応することを認知できる。キャラクターがユーザに対して一度振り向きを行い正対した後も、一定時間ユーザの発話音声の入力がない場合にはユーザから視線を外し、再び音声が入力されるとユーザに視線を戻す。以上の継続的な振り向き動作によって、ユーザの注意を維持することができる。

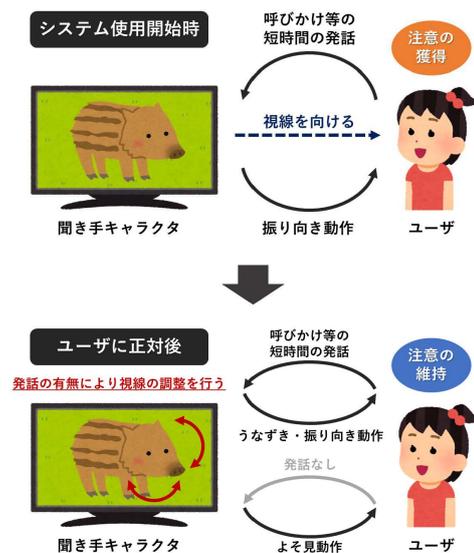


図1: 開発システムのコンセプト

2.2 予備実験

本研究では、システムを開発するにあたり、子どものユーザの注意の獲得、維持に対してキャラクタの振り向き動作が有効であるか検証を行うため、システムのプロトタイプによる予備実験を行った。

システムのプロトタイプでは、システム動作開始時、動物を模した3Dキャラクタが画面内の別オブジェクトへ視線を向けるように頭部を運動し続ける遊戯行動を行う(図2)。ユーザからの音声入力が行われると、



図 2: システムの使用画面

2.3 節のモデルに基づきモデルの予測値が閾値を超えた場合にキャラクタがユーザに対し振り向き動作を行い、ユーザを注視してから2.0秒経過後、再び遊戯行動を行う。ユーザを注視してから遊戯行動に至るまでの遷移時間は、Kendonの研究[4]の研究から、10人の持続時間の平均値が約2秒であったことから、先行研究[5]と同様に引用した。また、振り向き動作に対し、同モデルに基づきうなずき動作を行うモードとの比較を行った。これらのモードを小学生の実験参加者10人程度に提示し、システムに対し発話を促すような指示は出さず、自由にシステムを使用させた。システム使用後、システムを提示するモニタから離れた実験参加者に口頭でのアンケートを行った。アンケートで得られたシステムに対する感想に併せ、実験参加者がシステムを使用している際の行動から以下の点が明らかになった。

- うなずきのみを行うモードを使用した場合、常にユーザを注視するため、威圧感を感じてすぐにキャラクタへの対話を中止するユーザがいた。
- 遊戯行動が印象に残ったと答えるユーザが複数おり、遊戯行動がユーザの注意を獲得するのに有効である。
- 聞き手のインタラクションモデルに基づいた振り向き動作では、短時間の発話に反応しないことがある。
- 振り向き動作のみを行うモードの場合、初回の振り向き動作によってユーザの注意を獲得することはできるが、2回目以降の振り向き動作によってユーザへ継続的にかかわりを認知させることができず、ユーザの注意の維持が難しい。

以上の点から、システムの使用開始時にユーザの注意を獲得する手段として遊戯行動は有効であり、ユーザの発話音声に対し、キャラクタが反応を示すことを継

続的に提示することがユーザの注意の維持に重要である。そこで、短い発話時間でもユーザへの応答を行うよう、まず発話音声の初回入力時にユーザに対する振り向き動作を行い、その後はユーザと対面しインタラクションモデルに基づいたうなずき動作を行う中で、発話状況に応じてユーザを注視する時間を制御することにより、ユーザの注意を維持できると考えられる。

2.3 聞き手のインタラクションモデル

先行研究において、人同士の対面コミュニケーション時の発話音声と身体動作の関係を分析した結果、話し手の発話音声と聞き手の身体動作では、うなずきを主体とする頭部の動きとの相関が最も高いことが確認されている[3]。その解析結果に基づき、聞き手のインタラクションモデルとして、音声を閾値で二値化した場合のON-OFFパターンに基づくうなずき反応モデルが提案されている。うなずきの予測モデルはマクロ層とミクロ層からなる階層モデルである。マクロ層では音声の呼気段落区分でのON-OFF区間からなるユニット区間にうなずきの開始が存在するかを予測する。予測には、(2)式で表される $[i-1]$ ユニット以前のユニット時間率 $R(i)$ (ユニット区間でのON区間の占める割合)の線形結合で表される(1)式のMA(Moving-Average)モデルを用いる。

$$M_u(i) = \sum_{j=1}^J a(j) R(i-j) + u(i) \quad (1)$$

$$R(i) = \frac{T(i)}{T(i) + S(i)} \quad (2)$$

$a(j)$: 予測係数

$T(i)$: i 番目ユニットでのON区間

$S(i)$: i 番目ユニットでのOFF区間

$u(i)$: ノイズ

$$M(i) = \sum_{j=1}^K b(j) V(i-j) + w(i) \quad (3)$$

$b(j)$: 対象発話区間内でのON区間回数

$V(i)$: 音声データ

$w(i)$: ノイズ

予測値 $M_u(i)$ がある閾値を越えて、うなずきが存在すると予測された場合には、処理はミクロ層に移る。ミクロ層では音声のON-OFFデータ(30Hz, 60個)を入力とし、(3)式を用いてMAモデルでうなずきの開始時点点を推定する。予測値が閾値を越えた場合には聞き手キャラクタをうなずかせる。

2.4 開発システムの概要

本システムの開発は3DゲームエンジンのUnityで行った。システム使用開始後、聞き手キャラクタはシステム画面内の別オブジェクトに興味を示すような遊戯行動を行う。音声が入力されると、キャラクタはユーザの方向へ振り向き、体全体を回転させユーザに正対する。さらに、キャラクタがユーザに対して一度振り

向きを行い正対した後も、2.0秒以上音声の入力がない場合にはキャラクターは頭部を上へ15度、右へ15度回転させ、ユーザから視線を外す動作（よそ見）を行う。再び音声の入力が行われた場合には頭部を再び正対している際の姿勢になるよう回転させ、ユーザへ視線を向ける。ユーザに正対している間は、聞き手のインタラクションモデルに応じてキャラクターがうなずき動作を行う。開発システムのキャラクター動作とその遷移を図3に示す。

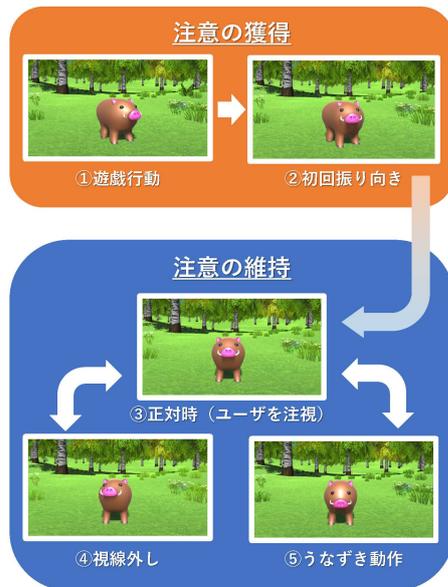


図3: キャラクターの動作と遷移

3 評価実験

3.1 評価実験の概要

本研究で開発したシステムを、2019年7月に開催された子どもたちを対象とした科学技術体験イベントにて展示した。この際、本システムの有効性を検証するために、本システムと比較対象のモードを来場者の小学生に使用させ、アンケートによる評価を行った。また、システム使用時の視線をキャリブレーションフリーの視線計測装置 (nac ACTUS) によって計測した。使用したモードを以下に示す。

背景 聞き手キャラクターのみ表示しない

A うなずき動作に加え、振り向き動作を行う

B ユーザに正対しうなずき動作のみを行う

背景モードは、実験の概要を実験参加者に説明する際に提示したモードで、このモードの使用時に視線を計測するセンサの動作調整を行った。Aモードが今回開発したシステムであり、Bモードは予備実験で使用した、従来のシステムと同様に一貫してユーザに正対するシステムである。背景モードを提示後、AモードとBモードの提示順が結果に影響しないよう、提示順はランダムとし、2つのシステムを一度ずつ実験参加者に使用させた。実験参加者は小学校3~6年生に在籍

する男女16人、うち3・4年生4人、5・6年生12人である。システムを使用する前に実験参加者には、学校に関する質問の提示を行い、これに対する回答について画面内に表示されるキャラクターに対し自由に話すよう指示した。また、実験参加者が話す内容がなくなり、発話をやめると判断した際には挙手をしてもらい、確認でき次第システムを停止した。

実験後、2つのモードを評価する一対比較と、キャラクターどの動作が好ましいと感じたか（うなずき動作、振り向き動作、その他）、システムを使用した感想の3つの項目の聞き取り式のアンケートを行った。また、システムの使用画面を図4のように分割し、システム使用時のユーザの視線の注視時間の割合と遷移を記録した。

なお、本研究は岡山県立大学倫理委員会の承認を受けており、実験参加者および保護者には同意書によってインフォームドコンセントをとっている。

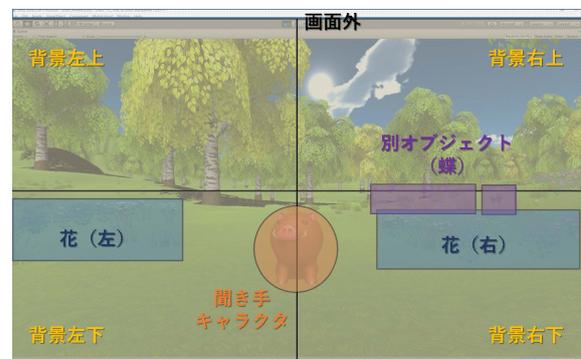


図4: 視線区分

3.2 官能評価結果

一対比較実験の結果を図5に示す。モードの使用順にかかわらず、振り向き動作を行うAモードが評価された。また、図6に動きの好みに関するアンケート結果を示す。アンケートでは選択肢に「うなずき」「振り向き」「その他」を用意したが、その他の動作が最も好ましいと答えた実験参加者はいなかった。一対比較でAモードがより好まれたが、一対比較でAモードを選択した実験参加者もうなずき動作がより好ましいと答える傾向がみられた。システムを使用した感想の中で、共通して挙げられた意見として以下のものが挙げられる。

- 振り向きの動作が良いと感じた。
- うなずきがあることで話しやすく、面白く感じた。
- うなずきに加えて、他の動き、表情や相槌が欲しいと思った。
- よそ見の動作が話を聞く意思がないことを示しているように見えた、視線をもっと合わせてほしい。
- キャラクターの見た目と動きが可愛かった。

これらの結果から、子どものユーザに対しても聞き手キャラクターがうなずき動作を行うことで、話しやすさを感じさせられることがわかる。一方で、うなずき動

作のみであると単調と感じるユーザが多く、相槌やうなずき以外の動きを求める声があり、振り向き動作が単調さを感じさせず、キャラクタの外見と相まって興味を獲得する傾向があることが示唆される。しかしながら、正対後に音声入力がない場合に行われるよそ見動作に関しては、話を聞く意思がないなど否定的な評価がみられることから、注意の維持に関して有効でない可能性が考えられる。

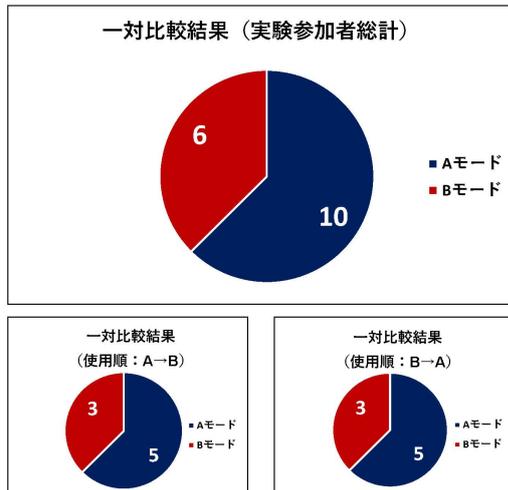


図 5: 一対比較結果

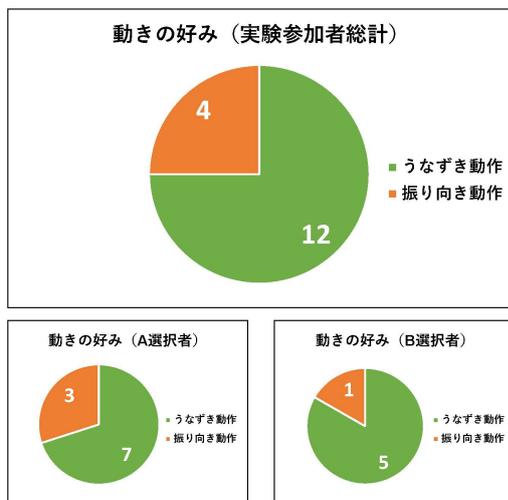


図 6: 動きの好みに関する結果

3.3 視線計測結果

実験参加者のうち、両モード使用時の視線情報を計測した 14 人について、各視線領域の注視割合について解析を行った。全体の傾向として、聞き手キャラクタやキャラクタが興味を示すオブジェクト（蝶）、蝶が位置する背景右側の注視率が他の領域に対し高い結果となった。両モード間で、領域別に注視率の一元配置分散分析を行った結果、全ての視線領域において有意差は確認されなかった。また、実験参加者のうち 5・6 年生 10 人の計測結果のみについて同様に一元配置分散分析を実施したが、実験参加者全体での結果と同様

に、有意差は確認されなかった。両モードにおいても、キャラクタではなく背景をより注視する傾向がみられる実験参加者が複数人いたことから、発話時に対象そのものを注視していなかった可能性が考えられる。

4 おわりに

本研究では、小学生以下の子どものユーザの発話促進、維持を目的とし、従来の iRT による聞き手の動作に加え、短時間の発話に対するキャラクタの振り向き動作により、ユーザの注意を獲得・維持するシステムを開発した。また、小学生を対象とした発話実験により、開発したシステムの有効性の検証を行うとともに、システム使用時の視線情報を計測した。評価実験における官能評価結果より、聞き手キャラクタがうなずき動作のみ行う場合の単調性を解消し、キャラクタの生物性を強調できている点で、振り向き動作による注意の獲得の可能性は示唆された。一方で、ユーザと正対時に意図的に視線を外す動作は、キャラクタ側からユーザとのかかわりを解消したと捉えられたことから、注意の維持には十分有効ではなかったと考えられる。評価実験において計測した視線情報の分析では、今回の解析条件においては両モード間で視線領域の注視率に差は確認されなかった。注視行動による発話時の行動分析についてさらに詳細な検討が必要である。

今後は、より子どものユーザが自発的に発話を行うようなキャラクタの動作の検討と、継続的な発話を実現する注意の維持を行うシステムの開発を予定している。

参考文献

- [1] 小吹健太郎, 上田博唯. ユーザの視線を感じて目を合わせる対話ロボットの提案. 研究報告コンピュータビジョンとイメージメディア (CVIM), Vol. 2012, No. 8, pp. 1-6, 2012.
- [2] 瀬島吉裕, 渡辺富夫, 神代充. 音声駆動型身体引き込みキャラクタに眼球動作モデルを付与した身体的コミュニケーションシステム (機械力学, 計測, 自動制御). 日本機械学会論文集 C 編, Vol. 76, No. 762, pp. 340-350, 2010.
- [3] T. Watanabe, R. Danbara, and M. Okubo. Effects of a speech-driven embodied interactive actor “interactor” on talker’s speech characteristics. In *The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. RO-MAN 2003.*, pp. 211-216. IEEE, 2003.
- [4] A. Kendon. Some functions of gaze-direction in social interaction. *Acta psychologica*, Vol. 26, pp. 22-63, 1967.
- [5] 深山篤, 大野健彦, 武川直樹, 澤木美奈子, 萩田紀博. 擬人化エージェントの印象操作のための視線制御方法. 情報処理学会論文誌, Vol. 43, No. 12, pp. 3596-3606, 2002.