

多人数仮想空間会話システムにおけるアバタ視線制御法の検討

宮島 俊光[†] 下地 崇[†] 藤田 欣也[†]

共有仮想空間を利用したマルチユーザ音声会話システムにおいて、発話情報を用いたアバタの視線制御法を提案する。提案手法では、各ユーザの発話の音圧情報に基づいてアピールポイントを算出し、アピールポイントが最大となるユーザのアバタを注視対象とした。主観評価実験をおこなった結果、自然なコミュニケーションへの有効性が確認された。

Gaze Control of Avatars in Multi-user Virtual Space Communication System

Toshimitsu Miyajima[†], Takashi Simoji[†] and Kinya Fujita[†]

In this study, we propose a gaze control algorithm of avatars based on the voice information in a multi-user virtual space communication system. Appeal point (AP) was defined, which is calculated using the voice level of each user. The gaze of each avatar was controlled with the AP. The effectiveness of the gaze control was demonstrated experimentally.

1. はじめに

ブロードバンドネットワークの普及に伴い、従来からの電子掲示板などのテキスト中心のものに加えて、音声や映像を相互に配信可能なメッセンジャーなど、より多様な情報を用いた遠隔コミュニケーションシステムが開発されている。さらに、グラフィクス描画機能の高速化により、仮想空間を用いたコミュニケーションシステムなども試みられている[1][2]。

CCD カメラを用いた実写型遠隔コミュニケーションシステムでは、顔の表情や動き、声の抑揚といったコミュニケーションに重要なノンバーバル情報を伝達することが可能であるが、カメラが必要であること、多くのネットワーク帯域を使用することに加え、実写画像を使用するため、背景画像によって意図しない個人情報伝送される可能性など、各種の問題が存在する。また同一の平面画面にユーザが表示されるマルチユーザ型システムでは、特定の相手との限定的な会話が困難であることに加え、全員の画像が平面的に並べられるため、誰が誰に向かって話しているのかわからず混乱をきたしやすい。そこで、会話者同士が視線を合わせる視線一致機能を、遠隔でも提供することが望まれている。

実写画像を用いた遠隔コミュニケーションシステムの上記のような問題点に対しては、さまざまな研究がなされている。例えば、プライバシー問題を解決する試みとして、実写映像ではなく3DCGを用い、さらにノンバーバル情報を伝えるために、カメラ映像を認識して顔の特徴点を抽出し、相手に表示されるアバタの表情を制御するシステムが開発されている[3]。これによりノンバーバル情報の伝達とプライバシーの問題は解決するが、視線一致のための制御法は別途考える必要がある。

視線の不一致を解決する試みとしては、ユーザ数に合わせて特殊なスクリーンを立体的に配置して、スクリーン背後にプロジェクタとカメラの両者を配置することで視線一致の問題を解決したシステム MAJIC[4]や、赤外線 LED と CCD カメラを用いてユーザの視線方向を検出できるアイトラックを利用して、仮想空間のアバタに反映させる GAZE[5]などが提案されている。さらに、音声認識によって会話対象を特定するシステムなども考えられるが、装置が大がかりでユーザ数が固定される点が難点である。特殊な装置を使用しないものとして、WWW ブラウザ上の動作が可能で、9方向の静止画像を用いて擬似的に視線の表示をおこなう BrowserMAJIC[6]もあるが、ユーザの注視対象をマウスポインタ位置で代替しているため、

[†]: 東京農工大学大学院

[†]: Tokyo University of Agriculture and Technology

各ユーザはマウス操作によって自らの注視対象を相手に伝える必要がある。

視線制御に関する研究とは別に、仮想空間内を移動することによって、会話相手を自由に選択可能なシステムも研究されている。FreeWalk[7]は、仮想空間内のウォークスルーと位置関係を用いた会話対象の制御を実現し、さらに知的エージェントを配置したときのユーザ挙動の解析が試みられているが、ユーザ間のコミュニケーションを促すための視線一致制御などの機能は有していなかった。

われわれも、仮想空間内の距離をもとに対話相手を制御し、会話開始を通知できる存在アウェアネス機能を実装した共有仮想空間歩行システムの開発を行った。しかし、会話に関係なく常にアバタの視線が一定方向を向くために不快感や違和感が生じ、コミュニケーションを阻害するという問題があった。本研究では、対話中における自然なコミュニケーションの促進を目的に、存在アウェアネス機能を有する共有仮想空間歩行システムをもとに、他のユーザの発話量から視線方向を制御する機能を実装し、その有効性について実験的に検討したので報告する。

2. システム設計

2.1 会話コミュニケーションシステム

それぞれ遠隔地にいる複数のユーザが仮想空間を共有し、共有仮想空間の中を自由に移動して相手を選択し会話できるシステムを実現するためには、位置や速度などの情報を相互に送受信し共有する仮想空間システムが必要である。

そこで、これまでに開発したクライアント/サーバ式多人数共有仮想空間歩行システム[8]をもとに、図1のようにP2P方式による多人数型音声通信機能を実装し、会話が可能なシステムを実現した。しかし、同一の仮想空間内にいるすべてのユーザの音声が届いた場合、現実空間では会話が困難なほど遠方のユーザとの会話が可能となり、違和感が生じる。そこで、アバタの周辺に会話可能圏であるオーラ領域を定義し、その領域内にいるユーザに対しのみ、会話が可能となる会話対象制御を行った。

さらに、オーラを用いた会話対象制御によ

て、自動的に会話相手を制御することが可能となるが、ユーザが会話可能であることを認識できなければ、円滑に会話を開始することができない。そこで、会話可能圏内に他者が入り会話可能になったことを自然に気づかせることが、自然なコミュニケーションの開始に重要な機能で、会話可能な対象の存在を自然に気づかせる、視覚と聴覚を介した存在アウェアネス情報提示機能を実装しその有効性を昨年度報告した。本研究では、この会話コミュニケーションシステムをもとに、視線制御機能を有する仮想空間コミュニケーションシステムを実現した。図2に使用時の風景を示す。

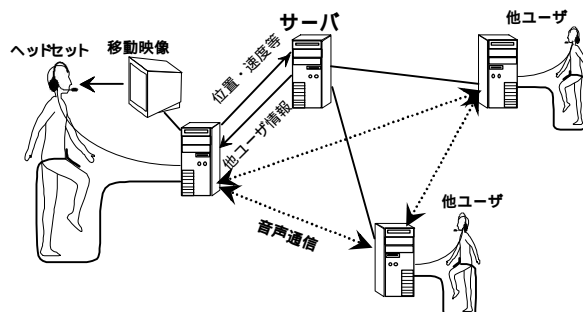


図1 遠隔共有仮想空間を用いたマルチユーザ歩行・音声会話システムの構成

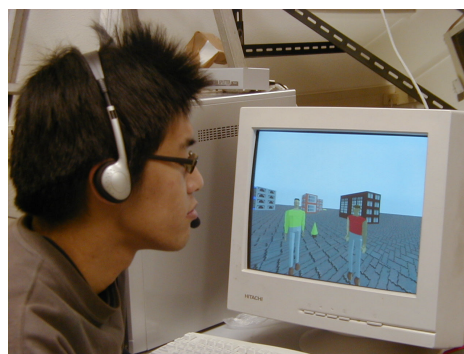


図2 仮想空間コミュニケーションシステムの使用風景

2.2 視線制御の実現

2.2.1 視線方向の制御方法

通常の現実空間における会話では、音声によって伝達されるバーバル情報だけでなく、ジェスチャや声の抑揚などのノンバーバル情報を利用することにより、様々な情報を相手に伝達している。その中でも、会話中の視線には、話しかける相手を同定したり、相手の話に耳を傾けていることを示したりする機能を果たして

いる。渡辺らの音圧情報をもとに「うなづき」動作を提示する研究[9]においては、会話の内容には関係なく音圧情報のみにもとづく制御であっても、大きな効果が得られることが知られている。本研究では、音圧を利用した各ユーザの発話情報と仮想空間における位置関係を利用して、それぞれのクライアントにおいて、擬似的にアバタの視線方向を制御する手法を提案する。

現実空間における多人数での会話を観察すると、聴衆は話者に対して視線を向ける傾向があり、また、会話中に他の一人が話し始めると、後から話し始めた方が注目を集める傾向がある。本研究では、前者を話者注視効果、後者を発言開始効果と呼び、これらの効果を注目度（Appeal Point 以下 AP）として定義し、音圧情報から算出される AP を用いてアバタの視線制御をおこなった。

2.2.2 話者注視効果(AP_c)

話者への注視を最も簡単に実現する計算方法は、各ユーザが発話しているか、していないかを検出し、現在発話しているユーザのアバタの方向に視線を制御する方法が考えられる。この場合、複数のユーザが発言した場合には、音圧の高いユーザの方向に視線を制御するのが妥当と考えられる。また、過去によく発言したユーザは、その後も発言する可能性が高いと考えられるため、現在だけでなく過去の発言量にも依存させるのが妥当と考えられる。これは、瞬時値に代えて、過去一定期間の発話総量を用いることで実現可能であるが、単純に適用すると、現在の発話中のユーザに視線が向かない可能性が生じる。そこで、話者注視効果 AP_c の算出にあたっては、現在の発話状況が優先される

ように、時刻 t における音量を $v(t)$ とするときの話者注視効果による注目度 AP_c は、(1)式の方法で算出し 0~1 に正規化した。

$$AP_c = \int_{-60}^0 \log(v(t)) \exp\left(\frac{t}{60}\right) dt \quad \dots (1)$$

過去 60 秒間の音圧の積分値を用いることで、他に発言者がいないときには、過去の発話量の多い者が注目を集める効果を実現した。また、現在値を優先するための重み係数は指数関数状に減少する時間関数とし、時定数 t は実験的

に 1.5 秒とした。計算に用いる音圧は、人間の知覚特性を考慮して対数変換を行った。

2 人のユーザが交互に発話したときを想定し、 AP の変化を図 3 に模式的に示す。区間のように、実際の会話でしばしば生じる発話の断続区間に対しても、話者注視効果 AP_c に積分を用いることで、即時に AP が 0 とならず、視線方向の頻繁な変化が回避可能となる。また、時定数を持って AP が漸減するようにしたことで、区間のように、発話が後も、当該ユーザがしばらくは注目を集めるという効果が実現される。

しかし、この方法を用いると、発話終了後すぐに AP が 0 にならないため、後から他者が発言を開始しても、区間の間は、先に発言していたユーザ 1 に視線が集まり、ユーザ 2 への注視が遅れて違和感が生じる。そこで、この問題を解決する方法として、次の発言開始効果 AP_s を導入した。

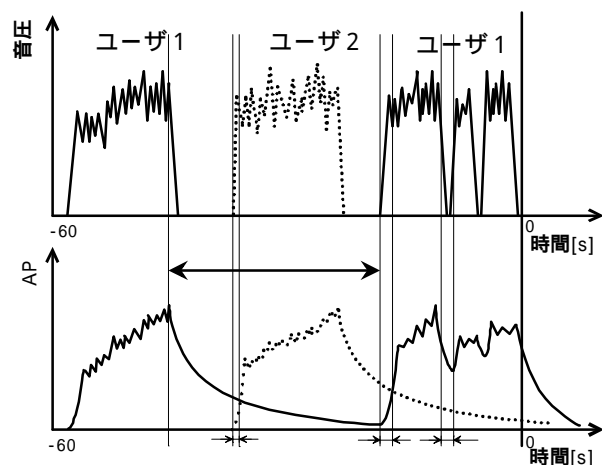


図 3 音圧(上)と AP(下)の関係の模式図

2.2.3 発言開始効果(AP_s)

発言開始効果は、上記の注視の遅れの問題に対処するために、発話の開始を検出して AP を高める処理である。音圧の過去の値が 0 で、現在値が 0 より十分に大きければ、発言の開始の判定は容易に可能である。しかし、実際の会話においては音声の断続が生じるため、発言開始と単なる音声の断続を区別するためには、発言開始検出後の一定時間は発言開始の検出を行わないなどの処理が必要である。また、発言開始時に注目を集める効果は、話者注視効果と比

較して瞬時的な性格をもつため、数秒以内に効果がなくなるのが適当と考えられる。

そこで、発言開始効果による注目度 AP_s の計算においては、発言開始と音声の断続を区別するため、過去 5 秒以内に発言がある場合には、発言開始効果が生じないものとした。また、発言開始効果は瞬時的な性格のものであるため、5 秒後に効果が 0 となるよう線形に減少させた。発言開始時刻を t_s とすると、時刻 t における発言開始効果による注目度 AP_s は、(2) 式によって定義される。

$$AP_s = \frac{5 - (t - t_s)}{5} \quad \dots(2)$$

$$(t_s \leq t \leq t_s + 5 \text{ かつ } t_{s+1} - t_s \geq 5)$$

2.2.4 注目度 AP の計算

上記の 2 つの要素を含む総合 AP は、話者注視効果による注目度 AP_c と発言開始効果による注目度 AP_s を用いて、以下のように定義した。 a, b の値は実験的に調整し、1:2 とした。

$$AP = aAP_c + bAP_s \quad \dots(3)$$

ここで、最も注目度の高いユーザのアバタを注視する処理をおこなった場合、話者交代が起きなければ、そのユーザを注視し続ける。しかし、長時間の注視は違和感を与える可能性があるため、乱数を用いて 30 秒に 1 回程度の頻度でランダムに注視対象アバタを強制的に変更する処理を行った。

3. 評価実験

話者注視、発言開始効果、およびランダム性の、自然で円滑な会話への効果を比較することを目的に、順序尺度を用いた好ましさの主観評価を行った。

被験者は本学学生 10 名を 5 名ずつの 2 群に分割し、各 5 名に対して、図 4 のように仮想環境内で他の 4 人が見える位置に各ユーザを配置し、験者から指示されたテーマに従って 5 分間会話する課題を課した。表 1 に実験条件の組み合わせを示す。

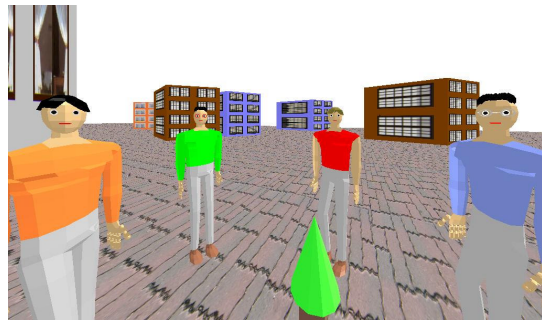


図 4 評価実験中の会話風景

表 1 視線アウェアネス制御評価実験の条件

なし
ランダム
発話量
発話量 + 発言開始効果
発話量 + ランダム
発話量 + 発言開始効果 + ランダム

実験で得られた順序尺度を、より好まれた条件で値が大きくなるように 0 から 5 点に点数化して平均したものを図 5 に示す。実験の結果、発言開始に対して小さな遅延でアバタの視線方向が変化する「発言開始効果」と、発話の音圧をもとに視線方向が変化する「発話量」を組み合わせた条件が、最も好まれる結果となった。

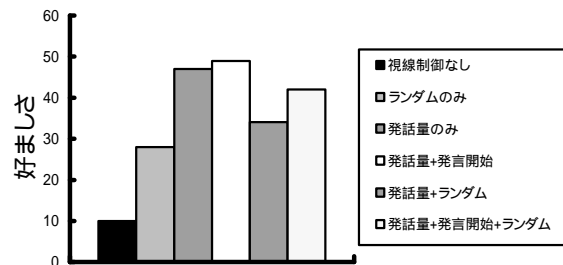


図 5 各視線一致条件の主観評価結果

実験で「視線制御なし」が最も好まれなかったのは、発話したにも関わらず、他ユーザの視線が全く変化しないため、視覚的に話者の特定が困難であったことが原因と考えられる。また、「ランダムのみ」が次に好まれなかったのは、各アバタの注視対象が異なるため、話者の特定に有効でないことによるものと考えられる。しかし、視線制御なしよりもランダムな視線制御が好まれたことから、アバタの視線方向が変化することによって、アバタに自律性が付与され、

このことが好まれたものとも推察される。

さらに、視線制御が好ましさの点で有効であった原因は、自らの発話によってアバタの視線が変化するインタラクティブ性、他のユーザの発話によって視線が変化する合理的な自律性、の2点が考えられる。

日常会話の場面において、話者は聞き手に注視されると相手が聞いていると認知するものと推察されるが、本システムでは発言開始効果を考慮して実装したため、ローカルユーザが発話しているときに他のユーザが発話を開始すると、そのユーザに他のアバタの視線を奪われ、相手が聞いている感覚が失われる、という問題が生じる。この問題を回避するためには、話者注視効果APcを他ユーザよりも大きくするなど処理が必要と考えられる。

また、仮想空間を利用した同時多人数コミュニケーションシステムでは、一度に多くのアバタを表示する必要がある。しかし、小型ディスプレイの限られた視野角では、人数が増えるとアバタの視認性が低下する、という問題が観察された。表情やコミュニケーションのための動作の実装にあたっては、検討が必要である。

4. まとめ

発話時の音圧情報をもとにアピールポイントを算出し、アバタの視線を制御するアルゴリズムを提案し、マルチユーザ共有仮想空間音声会話システムへの実装をおこなった。主観的評価実験を行ったところ、発話量と発言開始効果の組み合わせが、主観的に最も好まれ、自然な会話コミュニケーションに有効であった。今後の課題は、表情やジェスチャの制御機能を実装し、より自然なコミュニケーションを実現することである。

謝辞

本研究の一部は総務省戦略的情報通信研究開発推進制度によるものである、ここに記して感謝する。

参考文献

- [1] 松下, 岡田: コラボレーションとコミュニケーション, 共立出版(1995)
- [2] 垂水浩幸: グループウェアとその応用, ソフトウェアテクノロジーシリーズ第12巻, 共立出版(2000)
- [3] 沖電気, FaceCommunicator, <http://www.oki.com/jp/FSC/vc/fcbbe/about.html>
- [4] 前田, Giseok Jeong, 市川, 岡田, 松下, "MAJIC: 場の雰囲気重視したTV会議", 情報処理学会論文誌, 36.3, 775-783(1995)
- [5] Vertegaal, R. The GAZE GroupWare System: Mediating Joint Attention in Multiparty Communication and Collaboration. CHI'99, 294-301 (1999)
- [6] 岡田, 松下: 静止画像を用いた狭帯域ネットワーク用多地点会議システム, 情報処理学会論文誌, vol.39, no.10, 2762-2769 (1998)
- [7] 中西, 吉田, 西村, 石田: FreeWalk: 3次元仮想空間を用いた非形式的なコミュニケーションの支援, 情報処理学会論文誌, 39(5), 1356-1364(1998)
- [8] 下地, 藤田: 足踏式移動インタフェースWARPを用いた多人数共有仮想空間歩行システムの試作, 日本バーチャルリアリティ学会論文誌, 8(1), 11-18(2003)
- [9] 渡辺, 大久保, 中茂, 檀原: InterActorを用いた発話音声に基づく身体的インタラクションシステム, ヒューマンインタフェース学会論文誌 2(2), 21-29 (2000)