

仮想空間ボイスチャットシステムにおけるアバタの擬似視線制御

宮島俊光[†] 下地 崇[†] 藤田 欣也[†]

東京農工大学大学院[†]

1. はじめに

最近のネットワークの普及と高速化に伴い、テキストチャットや音声や映像を相互配信可能なメッセージャー、カメラを用いた遠隔会議システムなど、様々な遠隔コミュニケーションシステムが開発されている。

カメラを用いたマルチユーザ遠隔コミュニケーションシステムでは、全ユーザがカメラの方向を見るため、誰が誰に向かって話しているのかが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提供する試みとして、MAJIC[1]やGAZE[2]などが提案されている。しかし、装置が大がかりでユーザ数が固定される点が難点となる。特殊な装置を使用しない方法として、WWWブラウザ上で静止画像を用いて擬似的に視線の表示をおこなう BrowserMAJIC[3]が提案されているが、ユーザの注視対象をマウスポインタ位置で代用しているため、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。

他方、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の試みも行われている。筆者らも、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るという問題があった。本研究では、自然なコミュニケーションの促進を目的に、昨年度報告した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擬似的に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 仮想空間コミュニケーションシステム

現実空間で離れた場所にいる複数のユーザが仮想空間を共有し、その中を自由に移動することで相手を選択して会話することによってコミュニケーションを図るためには、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。

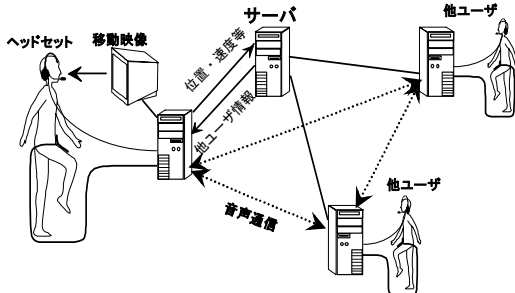


図1 遠隔共有仮想空間歩行・会話システムの構成

Pseudo control of avatar's gaze in shared virtual space voice chat system

[†] Graduate School, Tokyo University of Agriculture and Technology

本研究室では、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[4]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、会話可能なシステムを実現した。本研究では、このボイスチャットシステムをもとに、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。

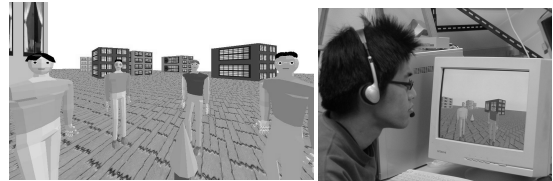


図2 評価実験中の会話風景

2.2 擬似視線制御の実現

2.2.1 視線方向の制御方法

実際の会話では、音声言語のバーバル情報だけでなく、ジェスチャや声の抑揚といった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[5]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、一般には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 (APc)

話者への注視を最も簡単に実現する方法としては、各ユーザの発話のオン/オフ状態を検出し、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 APc 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 APc は、(1)式の方法で算出し 0~1 に正規化した。

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

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 τ 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。

2.2.3 発言開始効果 (APs)

音圧の過去値が 0 で、現在値が 0 より十分に大きければ、発言の開始は基本的には検出できる。しかし、実際の会話においては音声の断続が生じるため、発言開始検出後の一定時間は発言開始の検出を行わないなど、発言開始と断続を区別する必要がある。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。

発言開始効果による注目度 AP_s については、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) 式で定義される。

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

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

2.2.4 注目度 AP の計算

話者注視効果と発言開始効果の両者を含む AP は、それぞれの注目度 AP_c と 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。

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

また、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、30 秒に 1 回程度の頻度でランダムに注視対象アバタを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、コミュニケーションの自然さに関する主観評価を行った。

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 2 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。実験条件として、話者の発話に依存しない視線制御なしとランダムのみとの 2 条件、および、話者の発話量を利用した話者注視効果をもとに、発言開始とランダムとを組合せた 4 条件を加え、計 6 条件で評価実験を行った。

被験者には、より自然なコミュニケーションが可能と感じた条件を、実験の最後に好ましい順に回答させ、順序尺度を求めた。実験の結果、他ユーザの視線が全く変化しない視線制御なしは、すべての被験者が最も不自然と感じ、次いで、発話状態と無関係に視線が変化するランダムのみが不自然な印象を与える結果となった。この 2

条件では、ランダムの方が自然な印象を与えたことから、 AP を用いた視線制御の 4 条件においても、ランダムを付加した 2 条件が、より自然な印象を与えるものと予想されたが、逆の傾向が見られた。複数のユーザがそれぞれ発話したため、視線制御をおこなう 4 条件においてランダム制御の有効性が期待される長時間の凝視が生じる場面が少なく、逆にアバタのランダムな視線移動が、話者注視効果や発言開始効果による視線制御を阻害したためと考えられる。結果として、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話者注視効果を組合せた条件が、最も自然な印象を与えた。

一般に、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、話者注視効果の AP を他ユーザよりも大きくするなどの処理が必要と考えられる。

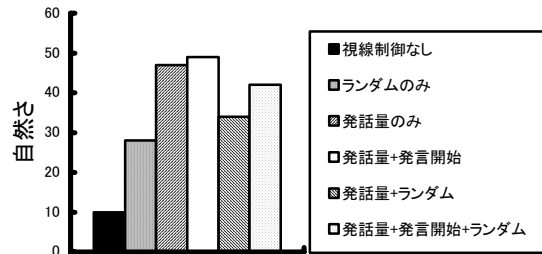


図 3 各視線一致条件における主観評価

4. まとめ

本研究では、ユーザの発話量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歩行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。

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]前田 他, 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)
- [2]Vertegaal, R. The GAZE GroupWare System: Mediating Joint Attention in Multiparty Communication and Collaboration. CHI'99, 294-301 (1999)
- [3]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)
- [4]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, 日本バーチャリアリティ学会論文誌, 8(1), 11-18(2003)
- [5]渡辺 他: 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)