

マルチユーザ音声チャットシステムにおける音声基本周波数を用いたアバタ表情制御法の提案

宮島俊光[†] 藤田 欣也[†]

東京農工大学大学院[†]

1. はじめに

高速インターネット網などインフラの整備に伴い、テキストチャット、電子掲示板、さらには音声や映像を伴うマルチメディアチャットシステムなど、様々なマルチユーザ型チャットシステムが開発されている[1].

実写画像を用いたビデオ会議システムは、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の伝達が可能となるが、実写画像を用いるため意図しない個人情報伝送への配慮が必要となる。これに対し、実写画像ではなく 3DCG を用いて仮想化することが有効と考えられ、現在、様々なアバタを用いた研究がなされている[2]. しかし、アバタを用いるシステムでは、表情や視線、身振りなどの身体動作を中心としたノンバーバル情報が欠如するため、これらの身体動作を適切に制御することが重要となる。

このノンバーバル情報を付加する試みとして、われわれも共有仮想空間マルチユーザ音声チャットシステムにおいて、音声情報をもとに注視対象の制御をおこなってきたが[3], 心的情報の表現が欠如する問題が生じていた。本研究では、仮想空間音声チャットシステムにおけるユーザの感情推定にもとづく表情制御方法について提案し、試作システムを用いて2名一組で会話をおこなわせたところ、提案手法は良好に被験者の状況を反映し、主観評価において有効性が確認されたので報告する。

2. システム設計

2.1 マルチユーザ音声チャットシステム

本研究室では、図1のようなサーバ/クライアント型の共有仮想空間ワークスルーシステム上に、マルチユーザ音声チャット機能を実装し、仮想空間内での位置関係を用いた会話対象制御をおこなってきた。さらに、各ユーザの音圧情報から Appeal Point を算出することで注視対象ユーザを決定し、カメラ等を用いずに音声情報のみから注視対象を制御する擬似視線制御を実装した[3].

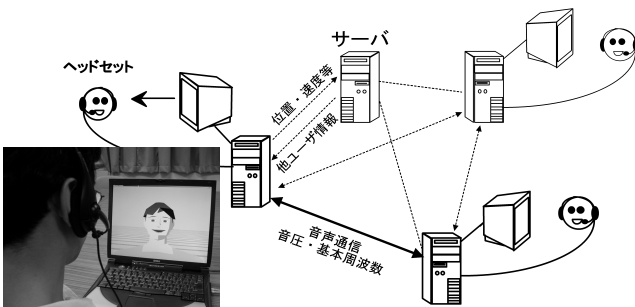


図1 実験風景(左)とシステム構成図(右)

Control of avatar's facial expression using fundamental frequency in Multi-user voice chat system

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

本研究では、図2のように、このような共有仮想空間マルチユーザ音声チャットシステムをもとに、音声情報を用いて表情制御をおこなう機能の実装をおこなった。

2.2 アバタ表情制御法

2.2.1 表情制御の概略

Russell は、感情を覚醒-睡眠と快-不快の2軸で表現する円環モデルを提唱している[4]. ここで、マルチユーザ音声チャットシステムへの応用を考えると、不快感情の表現は円滑なコミュニケーションが損なわれることが懸念されるため、快すなわち喜びの感情表現に限定した覚醒度のみの1軸の制御によってもコミュニケーションが円滑になる可能性が期待される。

本研究では、音声の基本周波数が喜びの感情を反映することを利用し[5], ユーザの音声情報から基本周波数をリアルタイムに算出して、その結果を受話側のユーザに音声とともに伝送し、受話側では受信した基本周波数にもとづいて喜びの感情値を算出してCGアバタの表情を自動的に制御する。

2.2.2 基本周波数の推定と音圧の算出

アバタのリップシンクや感情表現機能を実現するために音圧や表情制御のための基本周波数といった情報が必要となる。これら情報の取得、および音声通信には Microsoft 社製の DirectPlay 機能を用いた。また、基本周波数推定は、システムにおける処理負荷を考慮し、発話者の音声を発話側でおこない、ネットワークを介して他の受話ユーザへ伝送する形式を用いた。さらに、雑音の影響を回避するために、発話者の音声を 80ms 周期でバッファに取り出し、取り出した音声波形を3分割して、それぞれに対して区間 11ms の自己相関係数を算出した。3つの区間から自己相関値が最大となる周期を求め、値の近い2周期をもとに基本周波数を求めた。また、一定音圧以下の音声は雑音とみなした。

2.2.3 感情値の変換

基本周波数や感情によるその変化の大きさには個人差があるため、喜びの感情を適切にアバタに反映できるように、試作システムでは、平常時の基本周波数 F_{s0} と喜びの状態における基本周波数 F_{H0} をあらかじめ計測し、これを感情の最小値 0% と最大値 100% とみなし、線形変換した感情値 E_t の算出方法を、式(1)に示す。

$$E_t = \frac{F_t - F_{s0}}{F_{H0} - F_{s0}} \times 100 [\%] \quad (1)$$

また、発話内容による自然な抑揚によって、急激で不自然な表情の変化が生じるのを防ぐために移動平均処理をおこない、得られた感情値をもとに、CGアバタを描画した。笑顔画像は、FACS を参考に顔各部の代表点の移動をルール化し、プログラム中で自動生成した。図3に、アバタの笑顔の例を示す。

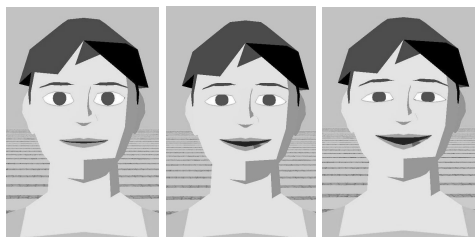
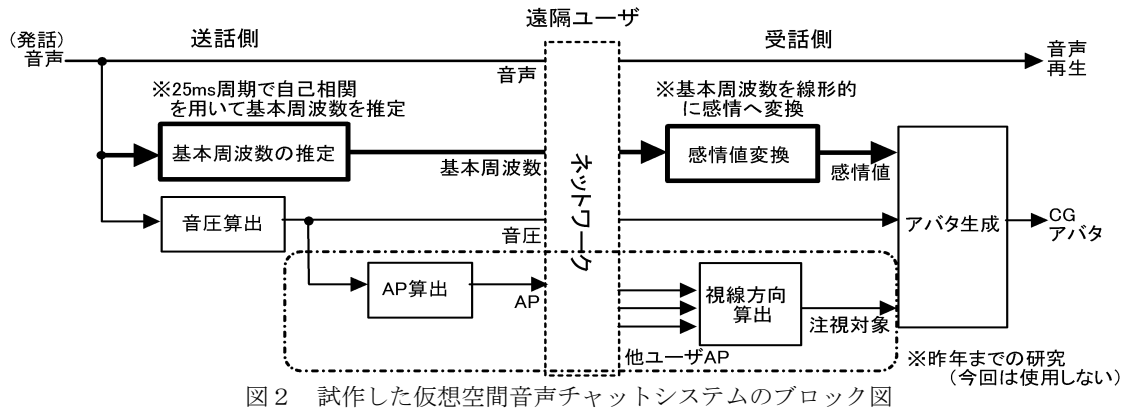


図3 アバタの笑顔 (左から 0%, 50%, 100%)

3. 評価実験

笑顔の提示、及び表情制御の効果を確認するため、表情が固定される平常顔(笑顔0%), 笑顔(笑顔100%), および話者の状態に応じて動的に変化する表情制御ありの3条件で予備実験をおこない、表情制御なしよりも、表情制御ありが主観的に良い印象を与えることを確認した。

つぎに、基本周波数を用いた表情制御において、不自然な印象につながる可能性がある急激な表情変化を抑制するための移動平均処理の適正値を検討することを目的に、順序法による好ましき、話しやすさ、および自然さに関する主観評価をおこなった。被験者は本学男子学生6名とし、2名ずつ3組に分け、被験者は別々の部屋でヘッドセットを着用し、仮想環境内の相手アバタを対面するように配置して、験者から指示されたテーマに従って4分間会話する課題を課した。

実験条件は、リップシンクのみ条件A、音声に対する表情の即応性を重視した移動平均区間2秒の条件Bと、表情の連続性を重視した移動平均区間10秒の条件Cの3条件とした。3条件を連続して会話させた後、主観評価とコメントについて質問紙を用いた聞き取り調査をおこない、順序効果を考慮して2回実施した。

順序尺度を用いた主観評価結果を図4に記す。移動平均区間2秒と10秒の条件BとCは、全ての主観評価において制御なしの条件Aよりも良い印象を与え、また、主観評価において条件BとCに顕著な差は見られなかったが、「条件Bは表情の移り変わりが速いため、感情が激しく変化しているように見えた」のコメントから、過度

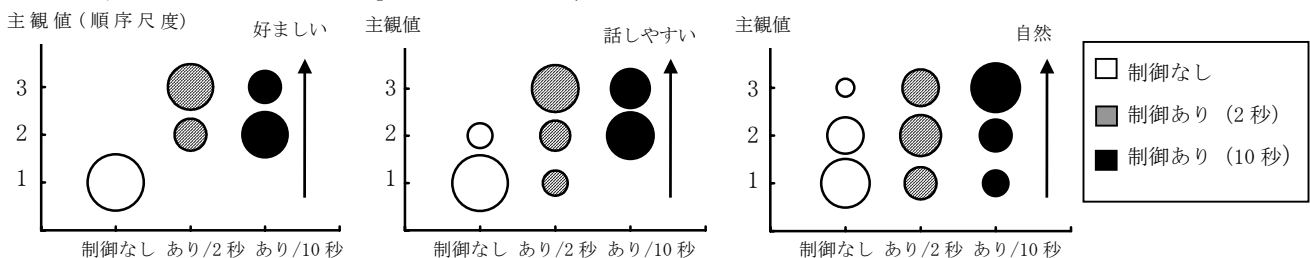


図4 実験結果 (円の大きさは回答数を意味する) ①好ましき (右) ②話しやすさ (中) ③自然さ (右)

な表情の変化はコミュニケーションに悪影響を与える可能性が懸念される。しかし、突発的な情動の表出には即応性が要求されるものと考えられることから、即応性と連続性の両者を満たすアルゴリズムの検討が今後の検討課題である。

4. まとめ

ユーザの発話音声をもとに基本周波数の推定をおこない、アバタの笑顔を制御するアルゴリズムを提案し、マルチユーザ音声チャットシステムへの実装をおこなった。評価実験をおこなったところ、表情制御アルゴリズムとして移動平均処理をおこなう2条件が良好な印象を与えたが、即応性と連続性の要求が背反することから2条件間で有意な差異は見られなかった。今後の課題は、表情制御アルゴリズムの更なる改良を加え、より自然なマルチユーザ音声チャットシステム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] 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)
- [3] 宮島, 下地, 藤田: 視線と存在の擬似アウェアネス機能を有する共有仮想空間コミュニケーションシステム, 日本バーチャルリアリティ学会論文誌, 10(1), 71-80(2005)
- [4] J.A.Russell: A circumplex model of affect, Journal of Personality and Social Psychology 39, 1161 - 1178 (1980)
- [5] 重永: 感情の判別分析から見た感情音声の特性, 電子情報通信学会論文誌, J83-A(6), 726-735(2000)