

# VR 音声チャットシステムにおける音声情報を利用した アバタ表情制御法の提案

宮島俊光<sup>†</sup> 藤田 欣也<sup>†</sup>

東京農工大学大学院<sup>†</sup>

## 1. はじめに

高速常時接続ネットワーク環境の普及に伴い、映像などを伴う様々なマルチユーザチャットシステムが開発されている。中でも、3DCG アバタを用いたシステムは、カメラ画像による意図しない個人情報伝送の回避に有効と期待される。しかし、アバタを用いたシステムでは、表情や視線、身振りなどのノンバーバル情報を補う必要が生じる。そこで、アバタ音声チャットシステムにおける話者音声情報をもとに、渡辺ら<sup>[1]</sup>は、筆者らは視線の制御を<sup>[2]</sup>実現したが、表情など心的情報にかかわるノンバーバル情報の制御機能が欠如していた。

ここで、アバタチャットシステムにおける表情の影響を考えると、不快感情の表出は円滑なコミュニケーションを損なう可能性が懸念される。そこで、Russellの感情円環モデル<sup>[3]</sup>における快-不快感情軸を快感(笑顔)に固定し、さらに音声の基本周波数が喜びの感情強度を反映することを利用して覚醒-睡眠軸(笑顔の強度)のみを音声基本周波数で制御するシステムを試作した<sup>[4]</sup>。しかし、相手の冗談による突発的な笑いのような、急激な表情変化の制御が課題であった。本研究では、発話開始時の音圧を併用することで、アバタ表情の瞬時的な変化を実現し、良好な結果を得たので報告する。

## 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

試作システムは、既報の共有仮想空間マルチユーザ音声チャットシステム<sup>[2]</sup>をもとに、VoIP機能モジュールに基本周波数と音圧の算出、ならびに通信機能を実装し、CGアバタの表情に反映させた。

### 2.2 アバタ表情制御法

#### 2.2.1 表情制御の概略

アバタの表情制御法の概略を図1に示す。マルチユーザ型のシステムにおいて、各ユーザに対して音圧や基本周波数をリアルタイムに推定することは処理負荷の増大を招くことから、発話側で話者音声の基本周波数および音圧を算出し、その結果を受話側のユーザに音声とともに伝送する方式を採用した。基本周波数は、区間11msの自己相関係数をもとに80ms毎の基本周期を求め、雑音や高調波の影響を回避するための、孤立値除去などの付加的处理を施した後に算出した。また、安定した基本周波数推定のために、性別の上限値(男性250Hz)を設定した。

受話側では、受信した基本周波数と音圧にもとづいて笑顔の強度を算出し、3DCGアバタの表情を制御した。

#### 2.2.2 連続成分SLcの算出

音声の基本周波数は快感によって高くなることから、ムードやゆっくりとした情動の変化による笑顔の連続的な変化成分SLc(Smile Level Continuous component)を、昨年の報告<sup>[4]</sup>と同様に、話者音声の基本周波数から算出した。

ただし、基本周波数の高さや、快感による基本周波数の変化量には個人差があるため、快感を適切にアバタに反映できるように、試作システムでは、(1)のように、平常時の基本周波数 $F_{S0}$ と、快感における基本周波数 $F_{H0}$ をあらかじめ計測し、これを笑顔の最小強度0%と最大強度100%とみなして線形変換した。

$$SLc = \frac{F_t - F_{S0}}{F_{H0} - F_{S0}} \times 100 [\%] \quad \dots (1)$$

また、基本周波数をそのまま笑顔制御に使用した場合、発話の自然な抑揚によって急激な表情変化が生じ、違和感を与える可能性が懸念されるため、5秒間の移動平均処理をおこなった。

#### 2.2.3 瞬時成分SLrの算出

上記の連続成分SLcは移動平均処理をおこなうため、表情が緩やかに変化して自然な印象を与えるが、急激な反応に対してはアバタ表情の遅延が生じ、違和感を生じることがあった。

そこで、会話相手の発言によってもたらされる、瞬時的な情動の変化を適切に笑顔に反映させることを目的に、話者交代時の発話開始音圧を、会話相手の発言によって誘発された反応の強さと見なし、笑顔の強度を自動的に制御する機能を実装した。試作システムでは、音圧の個人差等を考慮し、発話時の最小平均音圧 $V_{S0}$ と最大平均音圧 $V_{L0}$ をあらかじめ計測し、これを反応の強さの最小値0%と最大値100%とみなし、線形変換した。瞬時成分SLr(Smile Level Rapid component)の算出方法を式(2)に示す。

$$SLr = \frac{V_t - V_{S0}}{V_{L0} - V_{S0}} \times 100 [\%] \quad \dots (2)$$

ここで、SLrの算出には、他のユーザの発話に対する反応であるか否かの判定が必要であるため、他のユーザの発話が終了して1秒以内の発話を、他者への反応と見なし、SLrを算出した。また、他者の発話に対する急激な反応の多くは、瞬時的な一過性のものであるため、簡単な予備実験をおこなって、時定数1.5秒で指数的に減衰するようにした。

#### 2.2.4 総笑顔強度SLの算出と描画

算出された連続成分SLcと瞬時成分SLrを用いて、試作システムでは、式(3)で示すような重み付き線形和として笑顔強度SLを算出した。ここで、実験的に $a=0.75$ 、 $b=0.5$ とし、SLの上限値は100%とした。

$$SL = a \times SLc + b \times SLr \quad \dots (3)$$

Control of avatar's facial expression using voice information in multiparty voice chat system

<sup>†</sup> Graduate School, Tokyo University of Agriculture and Technology

また、アバタの表情は、顔各部の代表点が感情値 SL に比例して運動するよう FACS[5]を参考にルール化し、プログラム中で自動生成した。図 2 に、アバタの笑顔の例を示す。

### 3. 評価実験

各表情制御方法の効果の比較を目的に、5段階のリッカート法による主観評価をおこなった。実験条件は、表情制御なし、活性度成分 SLc のみ、反応成分 SLr のみ、SLc+SLr の4条件とし、順序効果を考慮して実施した。被験者には各条件で会話させた後、主観評価について質問紙を用いた聞き取り調査を実施した。被験者は本学男子学生10名とし、2名ずつ5組に分け、仮想空間に相手アバタが対面するよう配置し、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。被験者には、あらかじめ会話可能な内容について聞き取り調査し、共通な内容に対して会話させた。なお、実験中は、相手アバタを注視するよう指示した。

主観評価結果を図3に示す。すべての評価項目において、表情制御なしよりも、表情制御をおこなう3条件が良好な印象を与える結果となった。

表情が急激に変化する瞬時成分 SLr と、緩やかに表情を変化させる連続成分 SLc を比較すると、対話感と被聴取感においては SLr が、逆に、自然さと好ましさにおいては SLc が好印象を与えていた。SLc は、会話相手の活性度によって変化する音声の基本周波数のゆっくりとした変化をアバタ表情に反映させるため、より自然な印象に貢献したが、表情変化の時定数が大きいため、対話感には貢献しなかったものと解釈される。また、逆に、SLr は会話相手の反応の強さを反映することで対話感を増強したが、会話が活性化しても発話開始時以外の表情を変化させないため、自然な印象につながらなかったものと考えられる。

連続成分 SLc と瞬時成分 SLr の両者を組み合わせた条件が、全評価項目で最も好まれる結果となったのは、両

者が相補的に作用したものと推測され、自然でインタラクティブなコミュニケーションのためには、いずれもが必要と考えられる。

### 4. まとめ

ユーザの基本周波数および音圧をもとに、アバタの笑顔を制御するアルゴリズムを提案し、音声チャットシステムへの実装をおこなった。主観的評価実験からアバタの表情に発話者の緩やかな感情を反映する連続成分 SLc と、相手の発話に対する反応の強さを反映する瞬時成分 SLr の両者を組み合わせた条件で最も良好な結果が得られた。今後の課題は、表情制御アルゴリズムに更なる改良を加え、より自然な音声チャットシステムを実現することである。

### 謝辞

本研究は文部科学省特別教育研究費ならびに科学研究費補助金によるものである。ここに記して感謝する。

### 参考文献

- [1] 渡辺ほか: 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)
- [2] 宮島ほか: 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, 日本バーチャリアリティ学会論文誌, 10(1), 71-80(2005)
- [3] J.A.Russell: A circumplex model of affect, Journal of Personality and Social Psychology 39, 1161 - 1178 (1980)
- [4] 宮島, 藤田: 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, 情報処理学会第68回全国大会, 1G-3 (2006)
- [5] P. Ekman and W. V. Friesen: Facial Action Coding System, Consulting Psychologists Press, (1978)

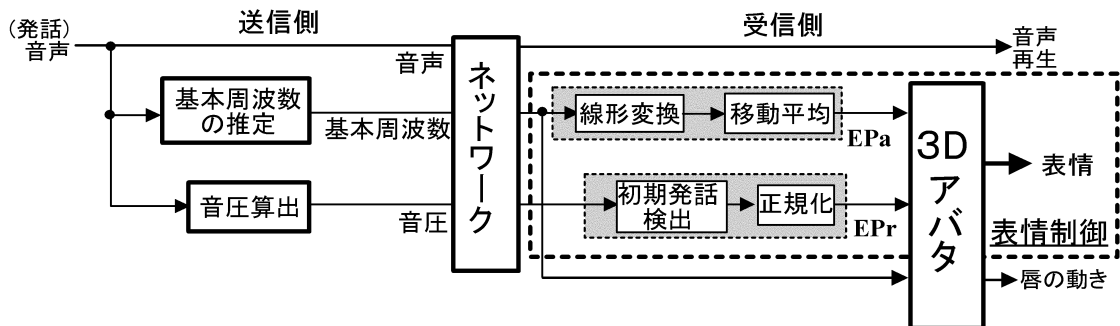


図1 システムのブロック図

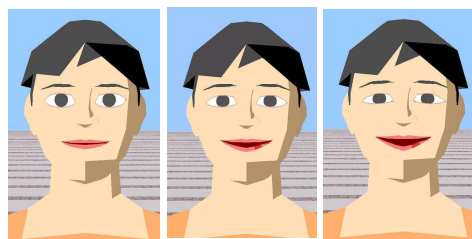


図2 アバタの笑顔 (左から 0%, 50%, 100%)

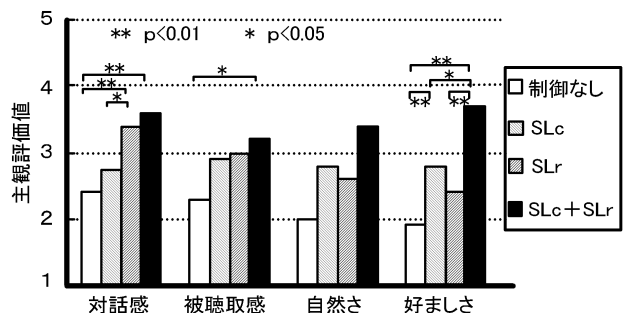


図3 主観評価結果