原著論文

Vol.9 No.4, 2007

# 音声チャットシステムにおける基本周波数と音圧を利用したアバタ表情制御法

宮島 俊光*1 藤田 欣也*1

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

Toshimitsu Miyajima*1 and Kinya Fujita*1

**Abstract** – **An automatic facial expression control algorithm of CG avatar based on the user's utterance is proposed, in order to facilitate multi-party casual chat in a virtual-space voice chat system. This study simplified the facial expression control problem by limiting the controling facial expression to the strength of the smile, because it is essential to facilitate casual chat. The proposed algorithm is comprised of three components those are calculated from the user and partner's utterance. The first is the smile level continuous component SLc that is defined as the moving-average of the normalized fundamental frequency of the partner's voice, to reflect the mood. The second is the smile level rapid component SLr that is calculated from the initial voice volume of the partner, to reflect the rapid emotion change caused by the user's utterance. The third is the smile level pseudo intention expression component SLpi that is calculated from the final phase voice volume of the user, to simulate the social response. A prototype virtual-space voice chat system that controls the avatar's smile strength using SL was developed. The effectiveness of the proposed algorithm and each component was experimentally demonstrated.**

**Keywords** : **Communication, Shared Virtual Space, Facial Expression, Avatar, Voice Chat**

## 1. はじめに

ネットワークの普及により，電子掲示板，テキストチャットなど，様々なマルチユーザ型チャットシステムが利用されている．特に，ネットワークの広帯域化を背景に，近年では，音声や映像を伴うマルチメディア型のマルチユーザコミュニケーションシステムが普及しつつある[1],[2].

実写映像を用いたビデオ会議型のチャットシステムは，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を伝達することが可能であるが，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意図しない個人情報伝送への配慮が必要となる．この個人情報伝送の問題に対しては，実写画像ではなくCGアバタを用いて仮想化する方法が考えられ，グラフィックエンジンの高速化を背景に様々なアバタチャットシステムの研究がなされている[3],[4],[5]．しかし，実時間実写画像の代わりに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を，手動操作などの何らかの方法で適切に補う必要が生じる．

筆者らも，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[6]，表情など他のノンバーバル情報を介した心的情報の表出機能が不足していた．表情制御に関しては，感情種が表示されたアイコンをマウスで直接操作し，相手アバタの表情を意図的に制御する研究[7]や，カメラ映像から顔の特徴点を抽出し，相手に表示されるアバタの表情を自動制御するFaceCommunicator[8]などがある．ほかには，FACS[9]に代表される人間の感情と表情の関係を用いて，なんらかの手段で推定したユーザの感情からアバタの表情を間接的に制御する方法も考えられる．ここで，感情推定の方法には，基本周波数やパワーなどの音声周辺言語に基づく方法[15],[16]や，音声認識をおこなってテキスト化したものから感情を推定するなど自然言語処理に基づく方法[11]が考えられるが，現在のところ，共有仮想空間を用いたマルチユーザ音声チャットシステムにおける，ユーザ状態推定に基づく表情制御は，脳波の乱れを用いて動揺状態などを表現する試み[12]が見られる程度である．

Russellは，図1のように感情を覚醒－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[13]．ここで，マルチユーザ音声チャットシステムにおけるアバタ表情の自動制御を考えると，システムの誤動作

*1: 東京農工大学大学院
*1: Graduate School of Tokyo University of Agriculture and Technology

による意図しない不快感情の表出はコミュニケーションを阻害することが懸念される．そこで本研究では，快すなわち喜びの感情表現に限定して覚醒度のみの1軸の制御に単純化し，さらに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て[10]，ユーザの音声情報から基本周波数と音圧をリアルタイムに算出して，話者アバタの表情を自動的に制御する方法を提案する．

自動制御されたアバタの表情は，必ずしも会話相手の実際の表情と一致しなくとも，ユーザが会話相手の音声から期待する表情とアバタの表情の間に大きな矛盾がなければ，連続的でインタラクティブな変化によって自然な印象や対話感を増強し，会話を活性化することが期待される．そこで，システムを試作して会話実験をおこなったところ，提案手法の自然さや対話感に与える効果が実験的に確認されたので報告する．

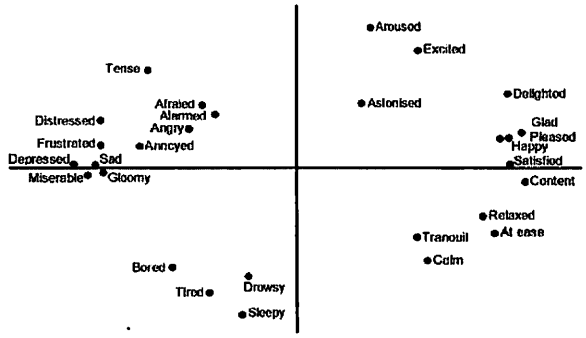


図1 Russellの円環モデル (水平:快ー不快，垂直:覚醒度)

Fig. 1 Circumplex model of affect (adapted from [13])

## 2. 感情と表情変化

### 2.1 感情モデルと感情認識

心理学の分野では，感情の分類に関する様々なモデルが提案されており，代表的なものとして，Ekmanが提唱する喜び・怒り・悲しみ・嫌悪・驚き・恐怖からなる基本6感情モデル[14]や，快不快と覚醒睡眠の2軸からなるRussellの円環モデル[13]などがある．

感情の分類やモデル化はいまだ心理学における議論の対象であるが，他方で音声や顔画像を用いた多くの感情推定の試みがなされている[15],[16],[17],[18]．顔画像や音声から得られる指標は，ある程度，感情種と相関のある結果を与えるが，例えば，音声では基本周波数が快感情だけでなく怒りでも高くなり，表情は一部の感情種間で相違が小さいなど，高い確度で複数種の感情を識別するのは，さらに研究を要する課題といえる．また，様々な感情種をアバタに表情として表現したときに，恐怖と驚きを誤認識しやすい[14]ことなども知られており，アバタに反映させるための認識を前提とすると，分離して認識する必要性が比較的低い組み合わせもあると考えられる．

ここで，表情自動制御機能を有するアバタ音声チャットシステムの用途を，友好的な状況でのカジュアルな会話に限定し，自然さや対話感の増強による会話の活性化をアバタ表情制御の目的とすると，システムの誤動作による意図しない不快感情の表出は，友好的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，不快感情の表出手段を手動制御に限定すれば，自動制御の対象は図1の円環モデルにおける快感情に限定されるため，覚醒-睡眠軸のみを考えれば良いことになる．すなわち，表情自動制御問題は，快感情による笑顔の強度を制御する1軸の制御問題に単純化して考える事が可能になる．

### 2.2 音声と感情

音声と感情の関係に着目すると，基本周波数は，喜びと怒りの感情で高くなり，他の感情では低くなる，という傾向がある[10]．ここで，先に述べた理由から怒りの感情を考慮しなければ，基本周波数が高くなる感情の変化は，喜びの快感情のみとなる．すなわち，怒りの感情を排除することによって，基本周波数の高さは，快感情の強さを反映すると見なすことができる．したがって，基本周波数を用いて笑顔の強さを制御することで，快感情が強い時には笑顔になり，悲しみなどの不快感情では無表情になる制御が実現するものと期待される．

### 2.3 笑顔の心理的要因

笑顔には多くの心理的要因があることが知られており，志水らは，笑いを快感情などによる本能的で不随意の笑いと，あいさつや追従などの社会的な随意の笑いの2つ，すなわち，感情表出の笑いと意思表出の笑いに分類している[19]．

また，人間の感情(Affect)を，ムード(Mood)と情動(Emotion)と2つに大別する考え方がある[23]．ムードは比較的長時間持続する穏やかな感情で，情動は認知や生理が関わる複雑な過程とされている．また，情動は血管収縮などの生体内部の変化と，顔をしかめるなどの観察可能な表出行動を誘発するとされている．

以上を考慮すると，音声チャットシステムにおけるアバタ笑顔の制御を考える場合，感情表出の笑顔と社会的な意思表出の笑顔の両者を実現することが望ましく，さらに，感情表出の笑顔に関しては，比較的穏やかなムードと，時には大笑いのように即時的な表出を伴う情動の，両者を実現することが，会話相手の音声から期待する笑顔に近い自然な表情変化につながるものと期待される．

これらを踏まえて，本研究では4章のように音声と笑顔の関係をモデル化し，アバタ音声チャットシステムに組み込んだ時の影響を実験的に検討する．

## 3. システム設計

### 3.1 マルチユーザ音声チャットシステム

筆者らは，サーバ／クライアント型の共有仮想空間ウォークスルーシステム[21]をもとに，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疑似視線制御法を提案し，その有効性を示した[6]．本研究では，このような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．

### 3.2 音声通信と音声情報の算出

アバタの口唇制御や感情表現機能を実現するためには，会話のための音声伝送に加えて，音圧や表情制御のための基本周波数の算出が必要となる．本研究では，音声通信と音圧の取得はMicrosoft社製のDirectPlay機能を用いて実装した．また，マルチユーザ型のシステムにおいて，各ユーザに対する基本周波数を受話側端末で推定することは，ユーザ数の増加につれて処理負荷が増大することから，発話側で発話者の基本周波数や音圧を算出し，ネットワークを介して受話側に伝送する形式を用いた．基本周波数の推定は，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．さらに，3つの区間から自己相関値が最大となる補周期を求め，値の近い2周期から基本周波数を求めた．このほか，安定した動作を得るために，簡単な予備実験をおこない推定周波数に250Hzの上限値を設け，一定音圧以下の音声は雑音とみなした．

また，音声の基本周波数には個人差があり，さらに直前の状態にも影響を受けやすい．そこで，個人の音圧と基本周波数の平均値，最低値および最低値を取得することを目的に，システムを使用する直前にキャリブレーションをおこなった．キャリブレーション値の取得方法は，「ももたろう」の最初の5文を，通常，大きい，小さい，高い，および，低い声で朗読させた．このとき，それぞれの発話時における平均音圧と平均基本周波数を算出し，この中で最も高い（低い）基本周波数を基本周波数の最高値（最低値）とし，同様に平均音圧の最高値（最低値）も記録した．

### 3.3 アバタの描画

チャットシステムで用いるCGアバタの表情は，次章に示す計算法を用いて算出される笑顔強度$SL$(Smile Level)を用いて，連続的に制御した．笑顔による表情変化は，FACS[9]を参考に，目，眉，口の代表点の移動パターンをあらかじめ定義しておき，笑顔強度$SL$に応じてプログラム中で代表点の座標を自動的に算出する方法によって制御した．図2に，アバタの笑顔の例を示す．

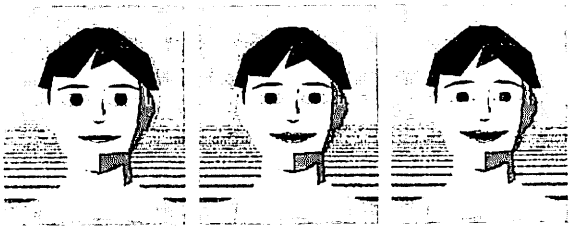

図2 アバタ笑顔（左から強度0%，50%，100%）
Fig. 2 Avatar's Facial Expressions ($SL$=0%, 50%, 100%)

## 4. アバタ表情制御法

### 4.1 表情制御の概略

先に述べたように，笑顔には主に快感情による感情表出の笑いと，社会的な意思表出の笑いの2つがある．さらに，感情表出の笑いには，比較的穏やかに変化するムードによるものと，突発的な大笑いのように瞬時的な情動によるものの2つがある．そこで本研究では，主にムードによる穏やかに変化する成分，情動による変化の中でも瞬時的に変化する成分，社会的意思表出による成分，の3つを近似的に音声情報から算出し，笑顔を制御する方法を考える．

パーティなどの場面で快感情によって会話が活性化してくると，徐々に話者の表情は笑顔へと変化し，このとき，音声の基本周波数は快感情によって徐々に高まる．したがって，話者音声の基本周波数のゆっくりした変化を検出して話者アバタの笑顔を制御すれば，ムードによるアバタ笑顔の変化を近似的に実現できる可能性が期待される．

情動行動は環境条件によって誘発され[20]，特に急激な笑顔の表出は，自己の内的要因よりも，他者の発話によって誘発される場合が多い．例えば，多者の冗談の内容が面白ければ誘発される笑い声や表情変化は大きく，面白くなければ声も表情変化も小さくなると考えられる．そこで本研究では，他者の発話に対する音声反応の強さを用いて，反応の強さに応じてアバタの笑顔を制御することで，瞬時的な情動による急激な笑顔の変化をモデル化することを考える．

また，受話者の実際の行動にはかかわらず，話者の発話終了を検出して，自動的に受話者アバタを頷かせる制御が引き込み効果を生み会話を促進したように[4]，追従などの社会的笑顔を自動的に提示することで，話者に対して自己の発話が聴取されている印象を与え，

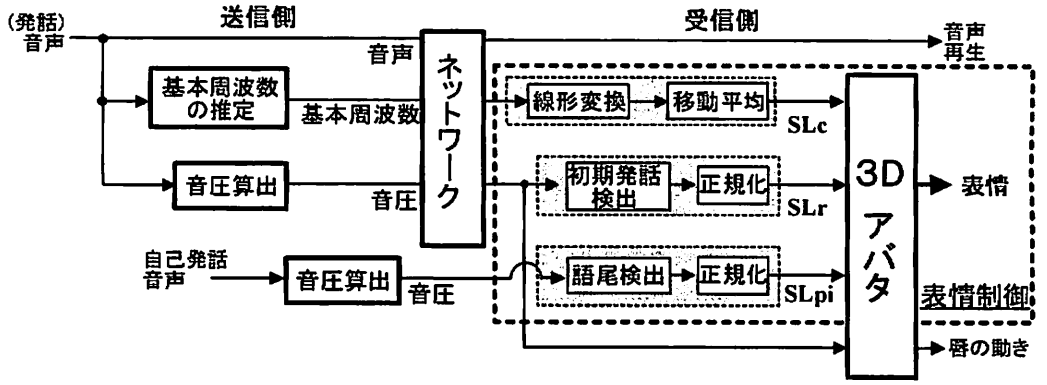


図 3 試作システムのブロック図
Fig. 3 Block diagram of prototype system

会話が促進されることが期待される．そこで，本研究では，受話者が音声による反応を示さない場合に，話者の発話終了を検出して，受話者アバタの笑顔強度を自動的に制御することで，疑似的に社会的意思表出の笑顔を提示するアルゴリズムを実装し，その影響を実験的に検討する．

試作したアバタ表情制御システム全体の構成を，図3に示す．発話側では話者音声の基本周波数を推定し，受話ユーザに音声データ，音圧とともに伝送する．受話側では受信した基本周波数と音圧に基づいて笑顔強度を算出し，発話者のＣＧアバタの表情を制御する．

### 4.2 連続成分 $SLc$ の算出

音声の基本周波数は快感情によって高くなることが知られていることから，本研究では，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，会話相手の発話音声の基本周波数から算出し，アバタに反映する．ただし，基本周波数の高さや，快感情による基本周波数の変化量には個人差があるため，快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，平常時の平均基本周波数 $F_{L0}$ と，快感情時の平均基本周波数 $F_{H0}$ 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．

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．そこで，基本周波数の移動平均処理をおこなった．図4は，被験者6名に「浦島太郎」の冒頭400文字程度を一定の調子で朗読させた場合の，基本周波数の1秒変動率の平均値を示す．抑揚のため，移動平均時間長1秒以下では50%以上の変動が見られるが，移動平均時間5秒では，変動率が20%以下に低減された．また，算出された値を用いてアバタの表情を制御し，被験者に聞き取り調査をおこなったところ，移動平均時間を5秒とすることで，ほぼ表情の変化が気にならないという内観報告が得られたので，本研究では，基本周波数の移動平均時間を5秒とした．$SLc$ の算出方法を式 (1) に示す．

$$SL_c = \frac{1}{5}\int_{-5}^{0} \frac{F_t - F_{L0}}{F_{H0} - F_{L0}} dt \cdot 100 \qquad (1)$$

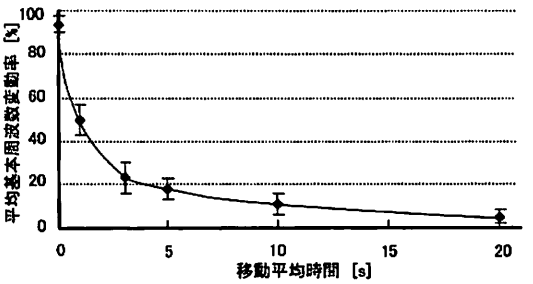


図 4 移動平均時間と平均基本周波数変動率の関係
Fig. 4 Relation between moving average time and deviation of averaged fundamental frequency

さらに，会話相手が非発話状態のときには，基本周波数を算出することができないため，なんらかの方法で $SLc$ を補間する必要がある．例えば，指数関数状に徐々に減衰させる方法などが考えられるが，実装したところ，ユーザの発話中にアバタの笑顔強度が減衰するため，会話内容に対して不安感を与えるとの内観報告があった．そこで本研究では，非発話状態における $SLc$ は，最終発話時に算出された値を保持するようにした．

### 4.3 瞬時成分 $SLr$ の算出

情動行動は環境条件によって誘発されるため，急激な笑顔は冗談などの他者の発話によって誘発される場合が多い．他方，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るため，違和感を生じることがある．したがって，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必要がある．そこで本研究では，話者交代を検出し，話者交代時の初期発話音圧 $V_{tp}$ を反応の強さと見なして，笑顔の強度を制御することとした．試作システムでは，マイク感度の影響や音圧の個人差を考慮し，発話時の最小平均音圧 $V_{L0}$ と最大平均音圧 $V_{M0}$ 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．

$$SL_r = \frac{V_{tp} - V_{L0}}{V_{M0} - V_{L0}} \cdot 100 \qquad (2)$$

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．

### 4.4 疑似意思表出成分 $SLpi$ の算出

$SLc$ および $SLr$ は，会話相手の音声から算出するため，会話相手が声を出さずに反応した場合には，アバタの表情は変化しない．そこで，本研究では，渡辺らが頷きの制御で用いた方法と同様にユーザの発話終了を検出し[4]，自動的に意思表出の社会的笑顔を提示する方法を考える．また，本研究では，単純に発話終了を検出して意思表出の笑顔を提示する方法に加えて，同意を求める発言や驚嘆の発声などは，語尾部分においても音圧が高いことが多いことから，語尾の音圧を話者の発話の強さとみなして，話者の発話の強さに比例して受話者の意思表出強度を制御する方法の，2つの制御機構を実装し，実験的に比較した．

提案手法では，音圧が，最小平均音圧 $V_{L0}$ の110%以下となった時刻を発話終了とみなし，その時点を基準に過去300msから700msにおける最大瞬間音圧 $V_{tu}$ を求める．最大瞬間音圧検出区間は，実験的に決定した．その後，3章で述べたように，あらかじめキャリブレーションで得られた自己の最小平均音圧 $V_{L0}$，および最大平均音圧 $V_{M0}$ を，話者の発話の強さに基づく疑似的な意思表出成分の最小値0%と最大値100%に相当するものとみなし，$V_{tu}$ を線形変換した．疑似意思表出成分 $SLpi$(Smile Level Pseudo Intention expression component) の算出方法を式 (3) に示す．

$$SL_{pi} = \frac{V_{tu} - V_{L0}}{V_{M0} - V_{L0}} \cdot 100 \qquad (3)$$

なお，追従の笑顔などの社会的意思表出は経時的に減衰して消滅するのが妥当と考えられるため，簡単な予備実験をおこなって，時定数 2.0 秒で指数的に減衰するようにした．

### 4.5 総笑顔強度 $SL$ の算出

算出された連続成分 $SLc$，瞬時成分 $SLr$ および疑似意思表出 $SLpi$ を用いて，試作システムでは，式 (4) のように，重み付き線形和として笑顔強度 $SL$ を算出した．

$$SL = a \cdot SL_c + Max(b \cdot SL_r, c \cdot SL_{pi}) \qquad (4)$$

ここで，$SLpi$ は会話相手の発声を伴わない意思表出を疑似的に実現する要素なので，発声を伴う応答から算出される瞬時成分 $SLr$ よりも小さい値を取り，また両者は同時に発生しないのが妥当と考えられる．そこで，$SL$ の算出には，荷重後の両者の最大値を用いた．重み係数は，表情制御に用いた時の違和感が無い値を実験的に決定し，a=0.75，b=0.5，c=0.25 とした．また，$SL$ は最大値が100%を越えないように制限した．

上記の算出方法によって算出した，会話中の $SLc$，$SLr$，$SLpi$（荷重前）および $SL$ の経時変化の例を図5に示す．好きな食べ物について2人に会話させ，徐々に会話が活性化し，53秒付近でユーザが言った冗談によって会話相手が笑った時のものである．

上段の基本周波数と連続成分 $SLc$ の関係を見ると，会話の活性化によって基本周波数が高くなり，$SLc$ も徐々に高くなる様子が読みとれる．また，基本周波数が抑揚で大きく変動しているのに比較して $SLc$ の変化は連続的であることから，移動平均時間はほぼ妥当と考えられる．

二段目の瞬時成分 $SLr$ に関しては，会話相手の発話初期音圧を反映し，その後指数関数状に減衰する様子が読みとれる．計測開始約5秒後の発話に対して算出されていないのは，ユーザの発話に対して1秒以内の発話をユーザ発話への応答とみなし，応答発話に限定して算出するアルゴリズムとしたためである．また，雑音によるユーザ発話の誤検出が約44秒で発生したため，多峰性の波形となっている．

三段目の疑似意思表出成分 $SLpi$ は，約30秒でユーザが発話終了した後に会話相手が応答していないため $SLr$ は算出されていないが，ユーザ自身の発話終了を検出して $SLpi$ が算出されるため，疑似的に意思表出の社会的笑顔が提示される．また，約35秒および44秒の二箇所で雑音による発話の誤判定が確認される．$SLr$ の算出にも影響するため，雑音による発話誤検出の軽減が必要である．

下段の総笑顔強度 $SL$ は上記の3要素の荷重和となっており，それぞれの要素を反映している様子が読みとれる．また，会話相手が音声を伴った反応をしない場合の疑似意思表出成分 $SLpi$ と，音声を伴った反応をした場合の瞬時成分 $SLr$ は，若干，疑似意思表出成分が先行するため，会話相手の発話開始において

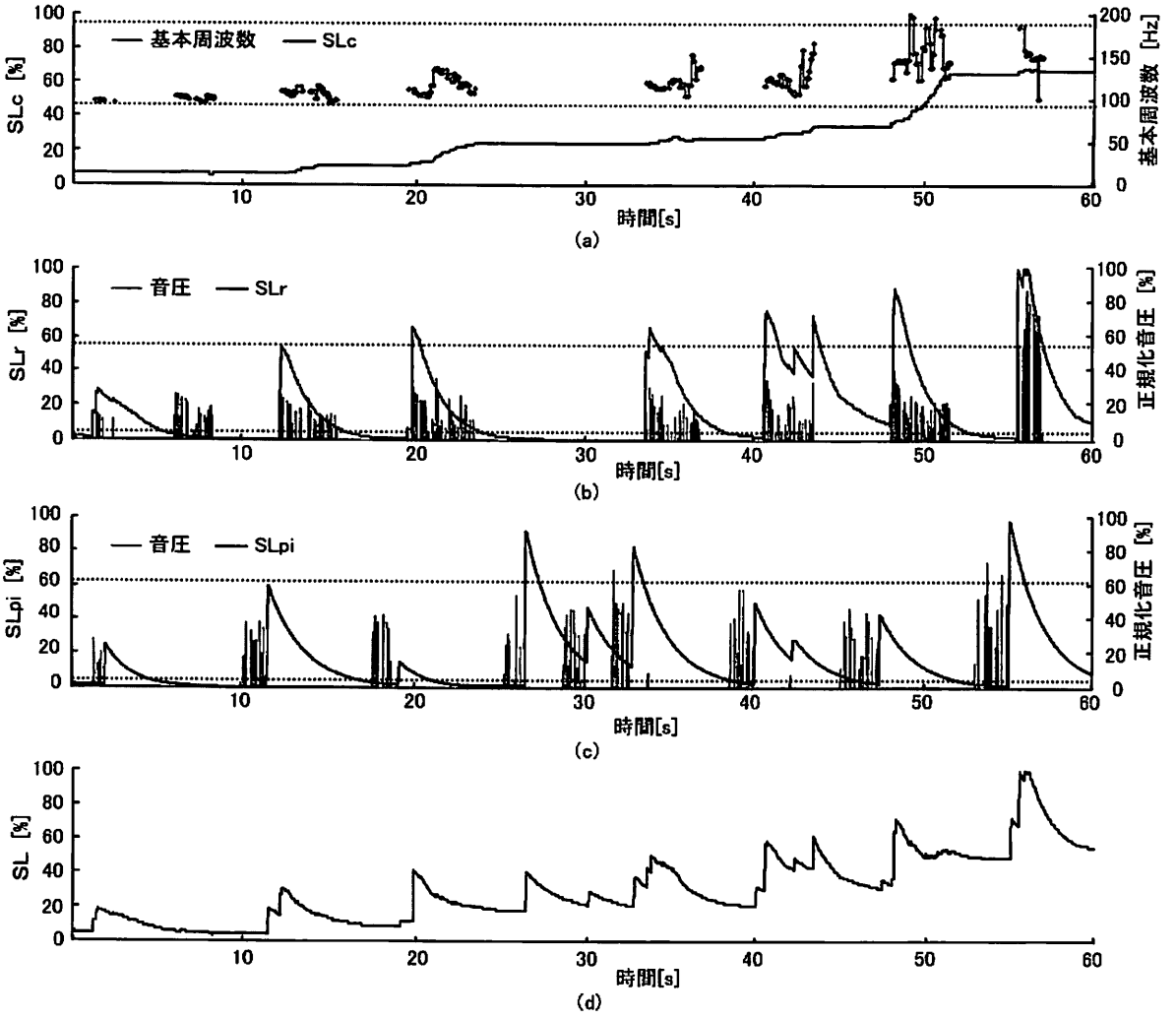


図5 会話時の笑顔強度 $SL$ の経時変化，(a) 会話相手音声の基本周波数と連続成分 $SLc$，(b) 会話相手音声の音圧と瞬時成分 $SLr$，(c) ユーザ音声の音圧と疑似意思表出成分 $SLpi$，(d) 総笑顔強度 $SL$

Fig. 5 Change of smile level $SL$ during conversation, (a)fundamental frequency of conversational partner and continuous component $SLc$, (b)voice pressure of conversational partner and rapid component $SLr$, (c)voice pressure of user and pseudo intention expression component $SLpi$, (d)total smile level $SL$.

軽度の二峰性を示しているが，実験的に予備評価を実施した範囲では，特に問題となる程度ではなかった．

## 5. 評価実験

### 5.1 連続成分を用いた会話実験

音声基本周波数を用いた表情制御の効果を検討することを目的に，連続成分 $SLc$ のみを用いて表情を制御し，リッカート法による主観評価をおこなった．被験者は男子大学生6名とし，2名ずつ3組に分け，図6のように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．課題として与える話題はあらかじめ日常的なものを選定し，さらにその中で被験者自身が長時間の会話が可能と考える話題を選別し，被験者同士で共通するものを実験時の話題として指定した．また，実験前にシステムを使用して自由に会話させ，環境に慣れさせた後に実験をおこなった．実験中は，会話相手のアバタを注視するよう教示した．

評価項目および具体的な質問内容を表1に示す．3条件を連続して会話させた後，主観評価と自由表記式の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った．また，順序効果を考慮して，以上の一連の実験を2回実施した．

### 5.2 連続成分と瞬時成分を用いた会話実験

対話相手の発話に応じて笑顔強度を算出する表情制御方法の効果を検討することを目的に，リッカート法を用いた5段階の主観評価をおこなった．実験条件は，表情制御なし，連続成分 $SLc$ のみ，瞬時成分 $SLr$ の

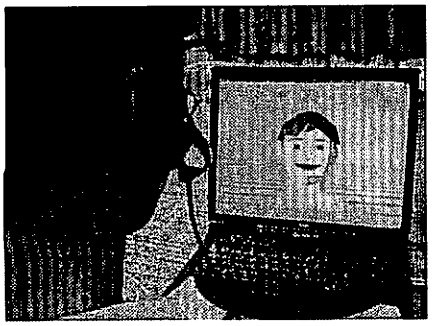

図6　会話実験風景
Fig. 6　Scene of conversation experiment

表1　評価項目と質問内容
Table 1　Evaluation items and questions

| | |
|---|---|
| 対話感 | 対話していると感じましたか |
| 被聴取感 | 聞いてもらっていると感じましたか |
| 自然さ | アバタの挙動を自然に感じましたか |
| 好ましさ | 対話をしていて好感を持ちましたか |
| 意思の反映 | アバタが会話相手の意思を反映していると感じましたか |

み，$SLc+SLr$ の4条件でおこなった．順序効果を考慮して各条件で会話させた後，前節と同様の評価項目に関して質問紙を用いた聞き取り調査を実施した．被験者は男子大学生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，実験開始前にキャリブレーションをおこない，実験終了までヘッドセットを固定し，さらに，実験中は相手アバタを注視するように指示した．

### 5.3　疑似意思表出成分を加えた会話実験

対話相手の発話反応に，自己発話を用いた疑似的な社交的笑顔を加えた場合の効果を検討することを目的に，リッカート法を用いた主観評価をおこなった．実験条件は，前節の実験で良好の結果が得られた $SLc+SLr$，$SLc+SLr$ に疑似意思表出成分を加えた $SLpi-Analog$，ならびに，疑似的な反応の有無の効果を見るために，$SLpi$ が10%以上なら100%に，10%より低ければ0%に変換した $SLpi-Digital$ の3条件で実施した．また，マルチユーザチャットシステムの利用場面として想定される中でも，発話応答しないユーザが発生して $SLpi$ の必要性がより大きくなる，3名会話条件で実験をおこなった．被験者は男子大学生18名とし，仮想空間に会話相手2名のアバタが左右に並んで表示されるよう配置して，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．また，順序効果を考慮し，1施行終了ごとに，表1の主観項目に関して聞き取り調査を実施した．それ以外の実験条件は前節と同様とした．

## 6.　結果

### 6.1　連続成分を用いた会話実験

連続成分 $SLc$ を用いた笑顔の制御に関する，主観評価結果を図7に示す．全評価項目において $SLc$ を用いて表情制御をおこなった条件が最も高い評価値となり，ほぼ全項目で有意差が確認された．常に一定の表情であるよりも，会話相手の状態に応じてアバタの表情が動的に変化することが，ユーザに自然さや対話感をもたらしたものと推察される．

また，表情強度固定条件では，笑顔100%条件の方が好まれると予想されたが，全項目で同程度の評価値であった．会話相手の状態を反映しない笑顔であるため，ユーザの好感につながらなかったものと推論される．

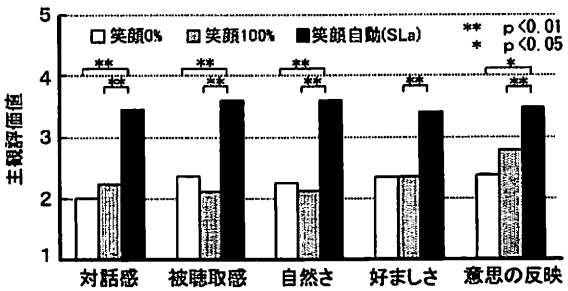


図7　表情制御条件と固定条件の主観評価結果
Fig. 7　Subjective evaluation of fundamental frequency based facial control and fixed face conditions

### 6.2　連続成分と瞬時成分を用いた会話実験

リッカート法による主観評価結果を図8に示す．すべての評価項目において，表情制御なしよりも，表情制御をおこなった3条件が良好な印象を与える結果となった．

表情が瞬時的に変化する瞬時成分 $SLr$ と，緩やかに表情を変化する連続成分 $SLc$ 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．$SLc$ と $SLr$ 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る．

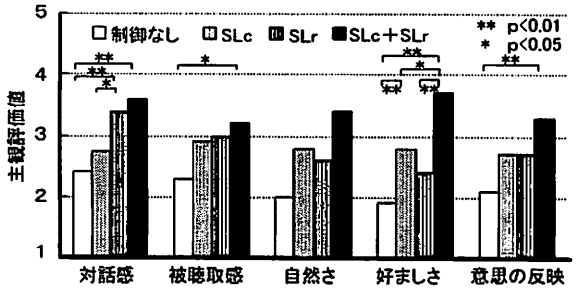


図8 連続成分と瞬時成分の主観評価結果
Fig. 8 Subjective evaluation of continuous and rapid components

### 6.3 疑似意思表出成分を加えた会話実験

リッカート法による主観評価結果を図9に示す．全体として，疑似意思表出成分 $SLpi$ がない条件よりも，疑似意思表出を加えた2条件でが好印象を与え，すべての項目において有意差が確認された．会話相手の，発話を伴わない表情による意思表出の疑似的実現が，対話感などにつながったものと見られる．疑似意思表出成分を加えた2条件に関しては，被聴取感を除いて両者の間に有意な差は認められなかった．

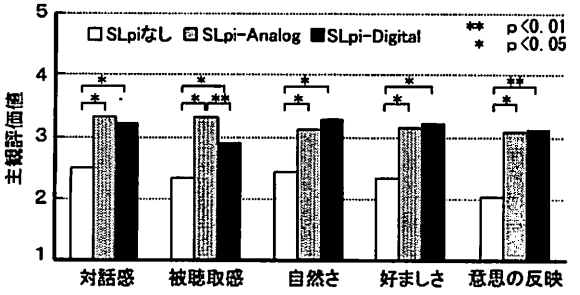


図9 疑似意思表出成分の主観評価結果
Fig. 9 Subjective evaluation of pseudointention expression component

## 7. 考察

連続成分を用いた会話実験では，表情を固定した条件よりも，会話相手の音声基本周波数を用いてアバタの笑顔強度を制御した方が，主観的に高い評価となった．伝送される情報の点では，音声基本周波数は会話音声から認知可能であるため，アバタ表情に反映しても会話相手に関する情報が増加する訳ではない．しかし，それでも好ましさや自然さの評価値が高くなった理由としては，一つには音声と表情の整合性が高くなったことが考えられる．さらには，同じ情報であっても，聴覚と視覚の複数のチャネルを介して提示されることで，相補的に強化された可能性なども考えられる．また，アバタの表示に音声基本周波数を反映することの妥当性に関しては，主観評価値が高くなったことから，本研究が想定する友好的状況のもとでの日常会話では，妥当であったと考えられる．しかし，対立状況での会話や，困難な交渉などの場面では，笑顔の強度制御のみでは問題となることが考えられ，カメラを併用して表現する表情種を増やすなどの方法を検討する必要がある．

連続成分と瞬時成分を用いた会話実験では，緩やかに表情を変化させる $SLc$ は，自然さや好ましさに，瞬時的に表情を変化させる瞬時成分 $SLr$ は，対話感や被聴取感に，より有効な傾向が見られた．瞬時成分 $SLr$ は，ユーザの発話に対して会話相手が発話によって反応した時にのみ算出される値であるため，$SLc$ と同様に，伝送される情報量を増加させるわけではない．しかし，ユーザの発話に対してアバタが反応することで，ユーザと会話相手のインタラクションが強化され，対話感や被聴取感の評価値が高くなったものと解釈される．

連続成分や瞬時成分と異なり，疑似意思表出成分 $SLpi$ は，ユーザの発話に対して会話相手が発話によって反応しなくても，自律的に笑顔を提示するものであり，すなわち疑似的に社会的な笑いを実現しようとするものである．今回の評価実験で，被聴取感の評価値が高くなったことは，渡辺らが発話終了を利用して頷き動作を制御することによって，引き込み効果を生み会話を促進した[4]のと同様に，笑顔を自己発話への肯定と認知したものと推察される．また，疑似意思表出成分の実験において，$SLpi-Analog$ と $SLpi-Digital$ の評価に有意差が見られなかったことから，疑似反応制御においては，声量によってうなづきを制御した研究[22]で示唆されたように，量的制御よりも，反応の有無の方が影響が大きいものと推察される．

このほか，笑顔に関する心理学的知見としては，笑いの種類による目と口の動き始めの時間的差異なども報告されている[23]．本システムでは，快の笑いと社会的な笑いの区別をしていないため，今後は表情変化パターンに種類を設けることなども検討が必要である．また，より自然で円滑なコミュニケーションシステムの実現のためには，表情と同時に表出されるジェスチャを中心とした，ノンバーバル情報提示機能の実装が望まれる．さらに，今回の実験システムは，比較的単純なCGアバタを使用し，アバタには会話相手の個性を反映していない．そのため，3人以上の複数人会話において発話者とアバタの対応がわかりにくいなど，より写実的なアバタの実現と併せて今後の検討課題である．

また，今回の実験では，自然さや対話感などの主観量への表情自動制御の効果が認められたが，それらの主観的効果の会話活性化効果に関しては，有効性を証明するにいたっていない．視線制御や他のノンバーバ

ル情報の制御も含め，アバタ自動制御のコミュニケーション活性化効果に関しては，さらに今後の検討が必要である．

本研究では，いくつかのパラメータを実験的に決定したが，より安定した動作のためには，これらの最適値の検討が必要である．また，音声基本周波数には個人差があるため，あらかじめキャリブレーションが必要であることに加え，基本周波数は体調などによっても変化するので，頻繁なキャリブレーションを避けるためには，正規化する，あるいは基本周波数そのものでなく抑揚の大きさを利用するなど，より実用的でロバストな動作のためのパラメータの検討が必要である．さらに，アバタ表情に応用した例がないため本システムとの直接比較はできないが，今後は，感情種や感情強度を推定する先行研究[15],[16]のアバタ表情制御への応用可能性などを含め，より広く表情制御の方法を比較検討する必要がある．

## 8. まとめ

ユーザ発話音声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，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．主観的評価実験では，アバタの表情に会話相手の緩やかな感情を反映する連続成分 $SLc$ と，瞬時的な反応の強さを反映する瞬時成分 $SLr$，さらにユーザの発話から算出する疑似意思表出成分 $SLpi$ を組み合わせた条件で，好ましさや対話感などにおいて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に検討し，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

本研究の一部は，文部科学省科学研究費補助金および特別教育研究費共生情報工学研究推進経費によるものである．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: 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] 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，Vol.39, No.5, pp.1356-1364 (1998).
[4] 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，Vol.2, No.2, pp.21-29 (2000).
[5] 本居，江崎，森本，黒川: 動的表情を合成したバーチャルアバタの感情伝達に関する実験的評価，ヒューマンインタフェース学会第 8 回ノンバーバルインタフェース研究会論文集，pp.21-24 (2004).
[6] 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，Vol.10, No.1, pp.71-80 (2005).
[7] 楠見，小島，米田: 3 次元マルチユーザ仮想空間におけるコミュニケーション，表情アバターによる感情表出と理解，情報処理学会 65 回全国大会講演論文集，No.5, pp.439-442 (2003).
[8] 沖 電 気: FaceCommunicator, http://www.oki.com/jp/FSC/vc/
[9] P. Ekman.and W. V. Friesen: Facial Action Coding System, Consulting Psychologists Press (1978).
[10] 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，Vol.J83-A, No.6, pp.726-735 (2000).
[11] 風間，植野，渡部，河岡: 常識的感情判断と主体語理，FIT2002, pp.137-138 (2002).
[12] 福井，林，山本，重野，岡田: 脳波計を用いたアバタの表情変化手法，日本バーチャルリアリティ学会論文誌，Vol.11, No.2, pp.205-212 (2006).
[13] J.A.Russell: A circumplex model of affect, Journal of Personality and Social Pshchology, Vol.39, pp.1161-1178 (1980).
[14] P. Ekman.and W. V. Friesen: 表情分析入門，誠信書房 (1987).
[15] 森山，小沢: ファジィ制御を用いた音声における情緒性評価法，電子情報通信学会論文誌，Vol.J82-D-II, No.10, pp.1710-1720 (1999).
[16] 柴崎，光吉: 抑揚からの感情認識の評価，信学技報，Vol.TL2005, No.15, pp.45-50 (2005).
[17] 下田，國弘，吉川: 動的顔画像からのリアルタイム表情認識システムの試作，ヒューマンインタ フェース学会論文誌，Vol.1, No.2, pp.25-32 (1999).
[18] L. Y. Tian,, T. Kanade and F. J. Cohn: Recognizing action units for facial expression analysis, IEEE Trans. PAMI, Vol.23, pp.97-116 (2001).
[19] 志水，中村，角辻豊: 人はなぜ笑うのか—笑いの精神生理学，講談社 (1994).
[20] 土田，竹村: 感情と行動・認知・生理，誠真信書房 (1996).
[21] 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，Vol.8, No.1, pp.11-18 (2003).
[22] 大橋，重本，森本，黒川: ノンバーバルモードの声量による制御を導入したバーチャルアバタの評価，ヒューマンインタフェース学会第 8 回ノンバーバル研究会，pp.11-14 (2004).
[23] 西尾，小山: 目と口の動きの時間的差異に基づく笑いの分類基準，電子通信学会論文誌，Vol.J80-A, No.8, pp.1316-1318 (1997).

(2006 年 12 月 11 日受付，2007 年 5 月 14 日再受付)

## 著者紹介

**宮島 俊光**

1997 年早稲田大学教育学部理学科数学専修卒業．現在，東京農工大学大学院工学教育部情報工学専攻技術職員として，バーチャルリアリティ技術を利用したコミュニケーション支援の研究に従事．

**藤田 欣也** （正会員）

1988年慶應義塾大学大学院理工学研究科修了．相模工業大学，東北大学医学部，岩手大学を経て，現在東京農工大学大学院教授．遠隔共有仮想空間および力触覚や歩行感覚の提示，ならびに医用福祉工学に関する研究に従事．工学博士．