2005－EC－1（4）
2005／6／4

Flash を用いた 3D 顔画像合成によるコミュニケーションシステムの構築

宮下直也†，佐藤隆†，松浦宣彦†，仲西正††
E-mail: {miyashita.naoya, satou.takashi, matsuura.norihiko, nakanishi.tadashi}@lab.ntt.co.jp
日本電信電話株式会社　NTT サイバーソリューション研究所

■あらまし　我々は，現在 98%以上の Web ブラウザにビルトインされている Flash をプラットフォームとして，3D 顔画像合成が可能なチャットシステムを構築した．Flash には 3D のサポートがないため，ActionScript を用いて独自に 3D 描画エンジンを実装した．これにより，ユーザは入力したテキスト情報をもとに，自身のアバタの表情をより簡易に伝えることができる．さらに，このシステムが通常の PC において十分なパフォーマンスを得られることを確認した．本稿では，本チャットシステムの特徴について述べる．

A Communication System using 3D Facial Image Synthesis on Flash

Naoya Miyashita†, Takashi Sato†, Norihiko Matsuura†, Tadashi Nakanishi††
E-mail: {miyashita.naoya, satou.takashi, matsuura.norihiko, nakanishi.tadashi}@lab.ntt.co.jp
NTT Cyber Solutions Laboratories, NTT Corporation

■Abstract　We developed a chat system that can synthesize 3D facial images by using Flash platform that is built in 98% or more of web browsers. As 3D graphics APIs are not supported on Flash platform, we originally implemented a 3D drawing engine in ActionScript. As a result, the user can communicate the facial expression of own avatar to other members more easily based on input text information. In addition, we confirmed enough performance was achived on usual PC on this system. In this paper, we describe the features of this system.

1．はじめに

近年，インターネットを介したリアルタイムコミュニケーション手段のひとつとしてチャットが利用されている．その中でも，ユーザ間でお互いの感情や場の雰囲気を感じ取れるようにするために，アバタを用いるものが提案されている．テキスト情報に比べて相手に感情を伝えられやすく，TV 会議システムのような実映像に比べてプライバシが保護しやすいという利点がある．しかしながら，従来は専用クライアントソフトウェアを必要とすることが多く，手軽に使うのが難しかった．

そこで，我々は，現在 98%以上の Web ブラウザにビルトインされている Flash をプラットフォームとして，3D 顔画像合成が可能なチャットシステムを構築した．Flash[1]には 3D のサポートがないため，ActionScript を用いて独自に 3D 描画エンジンを実装した．

また，アバタやチャットルームに存在する小道具等を多地点間で同期させるために，Java Servlet により，独自に同期サーバを構築し，スケーラブルかつ，他のサービスとの連携が容易になるようにした．小道具にはインタラクティブな I/F に CyberCoaster[2]

†日本電信電話株式会社 NTT サイバーソリューション研究所
NTT Cyber Solutions Laboratories, NTT Corporation
††日本電信電話株式会社　第三部門
DepartmentⅢ, NTT Corporation

を用いることで，エンドユーザが自由にそれを操作できるようにしている．さらにアバタや小道具は，Flash でコンポーネント化を行っているので，オーサリングツールで操作や配置が簡単に行えるため，クリエイタの参入が容易である．

本稿では，構築したシステムの詳細と評価について報告する．

2. 従来システム

リアル世界での人と人とのコミュニケーションにおいては，非言語的な合図（nonverbal cue）が重要な役割を果たす．人々は相手の姿勢や表情やジェスチャ等に注意を払い，そこから多くのことを知る．Ekman[3]は有名な一連の研究で，怒り，悲しみ，幸せ，嫌悪などの普遍的に認められる感情に対応する表情が存在することを示した．仮想空間内での人間同士のコミュニケーションがうまくいくためには，感情の合図を容易に表現したり読み取ったりできることが大切である．

従来のチャットシステムは，テキストをベースにしたものが主流であった．現在でも数多くのテキストベースのチャットシステムが存在する．感情をより効果的に相手に伝えるという点では，テキストベースでは不十分であるのが現状である．なぜならば，相手の感情がどういう状態なのかをテキスト情報のみからでは，読み取りにくいからである．

一方で，MSN や Windows メッセンジャーに代表されるように，表情付きのアイコンや音声，ビデオチャット等もできるようなシステムも存在する．しかし，顔文字やアイコンなどのような簡易な画像を用いたものでは，テキストベースのものと同様，そこから感情を読み取ることは難しい．ビデオ＋音声チャットなら相手の感情を読み取るのは容易であるが，プライバシ保護の面で利用者にとっては障壁が高いと考えられる．

これらを解決するために，仮想空間内を利用者の分身であるアバタで，お互いに会話をすることができるシステムが数多く存在する．上記に挙げたシステムと比較して，現実世界を仮想空間内に作り出すことで，臨場感を体験できるという利点がある．興味深い研究として，対面，ビデオ会議，3D 空間内のアバタの3者のコミュニケーション方法における発言量を比較して，3D 空間内の方が対面やビデオ会議よりも利用者の発言量が平等になったという報告もある[4]．しかし，感情を容易に伝えられないアバタなどの表現は，利用者に欲求不満を与える．さらに，専用のクライアントソフトウェアを用いるため，利用者の導入障壁が高い事が挙げられる．

我々は，3D で顔表情を合成し，仮想空間内の小道具をダイレクトでマニピュレーションするために CyberCoaster を用いる事で，相手に自分の感情や場の雰囲気を簡単に伝える事ができないかと考えた．表情を3D で合成する利点としては，2D に比べ，より自然な表情を作り出す事ができる点である．また，映像を直感的かつインタラクティブに操作できる CyberCoaster で作成した小道具を多地点で同期させる事によって，場の雰囲気や，場の一体感を伝える事ができるのではないかと考えた．そこで，以下の4点に着目し，コミュニケーションシステムを構築した．

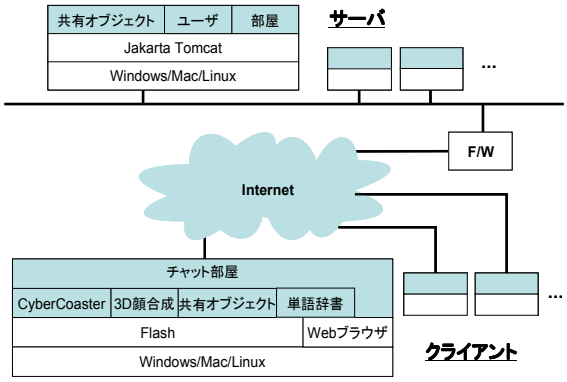
(1) Web ブラウザ以外の専用クライアントソフトウェアが不必要である事
(2) 通常の PC で軽快に動作する事
(3) 顔の表情や場の雰囲気を自然に伝えられる事
(4) クリエイタなどが参入しやすいオープンなシステムである事

次章では，構築したシステムの詳細について述べる．

3. システム構成

本システムの構成を【図1】に示す．サーバは，主にチャットルームに存在する利用者間で共有されるべきオブジェクトの管理を行う．例えば，アバタの表情変化や動きであったり，小道具の動きなどの同期制御を行う．その他には，ログインしたユーザの管理や部屋のテンプレート管理を担う．実装は，Java Servlet で行っている．クライアントは Web ブラウザにビルトインされている Flash で実装を行っている．サーバからダウンロードしたチャットルームに存在するオブジェクトの描画や利用者が操作したオブジェクトに対応するイベント通知をサーバに対して行う．次節以降でシステムの構成要素について，詳細を述

べる.



【図1】 システム構成

3．1． Flash による3D 描画エンジン

3D 顔画像合成を Flash で実装するにあたり，Flash は 3D Graphics API をサポートしていないため，独自に3D 描画エンジンを ActionScript で記述する必要がある. Flash では座標平面 MovieClip(以下，MC と略)それぞれについて，以下の幾何変換が可能である(【図 2】).

(1) X 軸, Y 軸方向の拡大･縮小.

拡大率はそれぞれ xscale と yscale

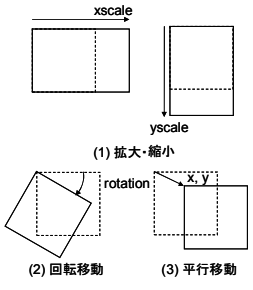
(2) 原点を中心とした回転移動.

回転角 rotation

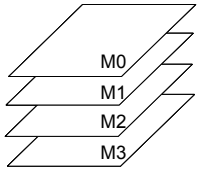
(3) 平行移動.

上位の MC に対する原点の位置(x, y)

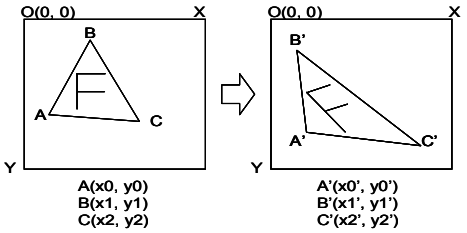
MC は【図 3】のような階層構造をもつ. 上記(1)～(3)の MC の変換パラメータを複雑に組み合わせることによって，【図 4】のように任意の三角形 ABC を任意の三角形 A'B'C'に変換するアフィン変換を行うことを考える.



【図2】 MCによる幾何変換

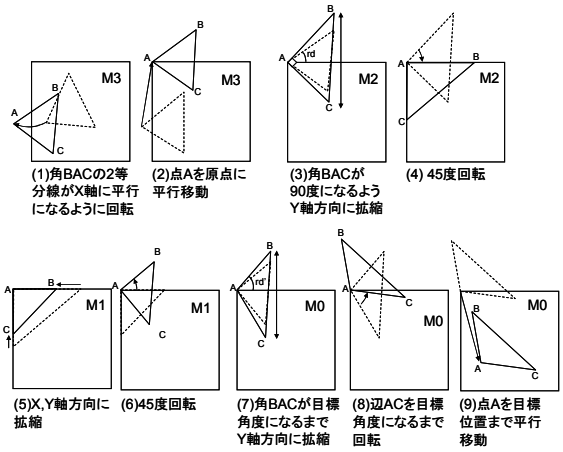


【図3】 MCの階層構造



【図 4】三角形パッチの幾何変換

【図 5】に， 各 MC のレイヤにおいて， 三角形 ABC がどのように変形されるかを示す. M3～M0 の表示は，その段階でパラメータを設定する MC を表す. 点線の三角形は，変形前の形状を表し，実線の三角形は，変形後の形状を表す.



【図 5】 アフィン変換手順

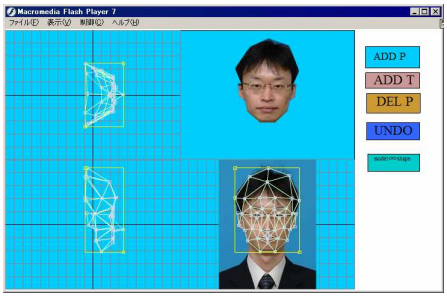
3．2． 3D 顔画像合成

人物の顔の表情や口の形状は，骨格の移動，顔面筋や皮膚の伸縮によってさまざまに変化する. これらの動きによる変化をパラメータ化し，そのパラメータに対する表情の変形規則が定義されれば，これらのパラメータの組み合わせによって表情合成が可能となる. この表情を記述するパラメータとして，心理学の分野で提案されている FACS(Facial Action Coding System)[5]を適用する手法が提案されている. また，口形状を記述するパラメータも提案されている[6].

本手法では，上記の表情パラメータ，口形パラメータの変形ルールを 3 次元ワイヤフレームモデルに適用する事で任意の表情を合成できるようにしている. 変形を行うための幾何変換は，3.1 で述べた Flash で構築した三角形パッチを任意の三角形

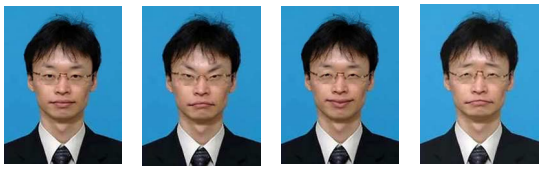
パッチに変形できる 3D 描画エンジンを用いる．【図 6】に Flash で構築した整合ツールにより，2 次元画像に 3D ワイヤフレームモデルを整合した結果を示す．今回，用いた 3D ワイヤフレームモデルの三角形ポリゴン数は約 80 ポリゴンである．【図 7】に表情を合成した結果を示す．少ないポリゴン数にもかかわらず，ある程度自然な表情合成が可能である．

本システムにおける，表情の駆動方法は，あらかじめ感情を表す単語を辞書に登録しておき，利用者が入力したテキストが登録単語に部分一致した場合に，対応する表情を合成するという手法



【図 6】2 次元画像の整合結果

をとっている．マウスやキーボードのファンクションキーによるダイレクトマニピュレーションに比べると，テキストによる会話の中で自然に表情を表出できる．現在の登録単語数は100語程度である．

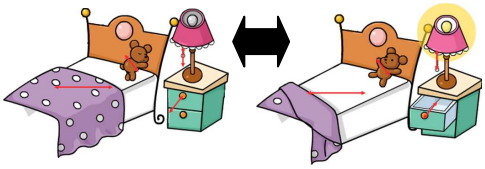


平静　　怒り　　幸せ　　悲しみ

【図 7】 表情合成結果

3．3． CyberCoaster

CyberCoaster とは，映像の中に映った人物や物体を，マウスを使ってあたかも直接つかんで動かしているように再生できるソフトウェアである．映像中に折れ線スライダを配置する事で，その折れ線に沿って直感的に物体をインタラクティブに動かすことが可能である．【図 8】にチャットルームに存在する小道具に対して CyberCoaster を適用した例を示す．スライダ（図に示す矢印）を画像中に配置する事で，線にそって直感的かつインタラクティブに小道具を動かす事ができる．例では，小熊のぬいぐるみ，電灯，かけ布団，引き出しをインタラクティブに動かした結果を示している．スライダは，表示/非表示の状態を自由に切り替えることができる．



【図 8】CyberCoaster の適用例

3．4． 同期サーバ

チャットルームに存在する小道具やアバタは，マウスやキーボードによるアクションに応じて，利用者間で同期を取らなければならない．つまり，ある小道具を，場を共有している利用者の一人がマウスで動かすと， その動きを他の利用者が見ているチャットルームに反映させることが必要となる． この処理を同期サーバが受け持つ．共有されるべきオブジェクトには，以下のものが存在する．

（1） アバタの表情・口形パラメータ
（2） CyberCoaster による小道具の位置パラメータ
（3） 利用者が入力したテキスト情報
（4） 利用者を識別するためのメタデータ

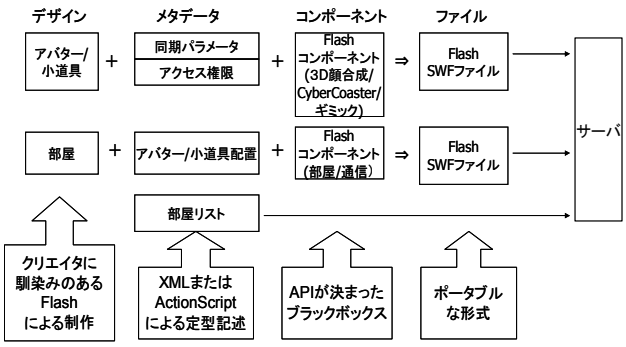
これらのオブジェクトは， Flash で構築したチャットルーム（クライアント）と同期サーバ間でシリアライズ化されたデータとして送受信される． チャットルームは， 利用者が，その中に存在するオブジェクトの状態を変更すると， 同期サーバに対して， 状態変更があったオブジェクトの情報をシリアライズ化して通知する． また，適当な間隔(10～20ms)で，サーバに対して，他の利用者のオブジェクトが更新されているかをポーリングする．同期サーバ側では， クライアントから受信したオブジェクトを他のチャットルームに通知する． これにより，利用者間でのオブジェ

クトの同期が実現できる．チャットルームの描画に関しては，すべてクライアント側で行っている．

同期サーバは，Jakarta Tomcat をコンテナとする Java Servlet で実装を行っている．また，企業や特定の ISP 等では F/W(ファイヤウォール)により，HTTP プロトコルの通信ポート(80 番)しか通過できないのが一般的である．よって，本同期サーバでは，HTTP プロトコルにより通信を行うよう工夫を行った．

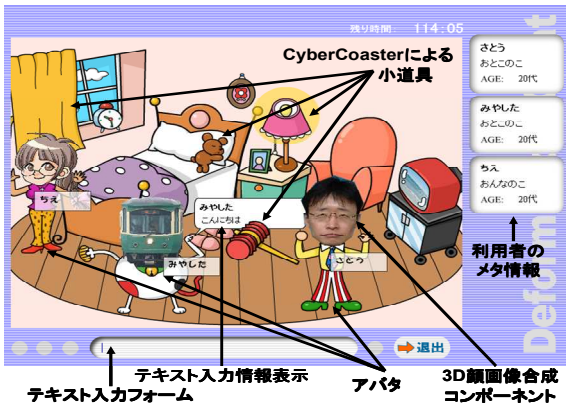
3．5．オープンコンテンツ構造

クライアント側は，クリエイタ等の導入障壁を下げるために，オープンなコンテンツ構造にしている【図 9】．クリエイタやシステムの運営者は，我々が実装した Flash による 3D 顔画像合成コンポーネント，CyberCoaster コンポーネント，部屋や通信を行うコンポーネントを利用して，アバタや小道具をオーサリングツールで製作し，部屋に配置し，最終的に Flash の SWF ファイルにパッケージングしてサーバにアップロードするだけで，簡易にチャットルームを開設する事ができるようになっている．



【図 9】クライアント側のソフトウェア構造

【図 10】に，本章で説明した各要素を統合して構築したチャットルーム(クライアント)の一例を示す．チャットルームは，上記で説明したようにクリエイタが自由に定義できるオープンな構造となっているので，短期間で部屋のデザインを変更したり，時事的な話題を反映した部屋を定義したりする事も簡易にできるという利点がある．



【図 10】チャットルーム

4．評価

本システムの評価方法は，チャットルームに存在するアバタ数を増加させていった時のクライアントシステム側の描画速度とその時の CPU 使用率，オブジェクトがネットワークを流れるトラフィック量，さらに同期サーバ側のパフォーマンスを計測する事で総合的に十分なパフォーマンスが得られるかを確認する．実験環境は以下に示すとおりである．

クライアントは，

OS:WindowsXP Pro

CPU:Pentium4－1.6GHz×1

Memory:1GB

グラフィックカード:ATI MOBILITY RADEON 7500　32MB

ブラウザ:IE6.0 SP2(FlashPlayer7.0)

ワイヤフレームのポリゴン数:80

アバタの顔に用いる画像サイズ:100x100(pixel)

同期サーバは，

OS:RedHat Linux9.0

CPU:Pentium4－1.9GHz×1

Memory:512MB

サーブレットコンテナ:Jakarta Tomcat5.5.

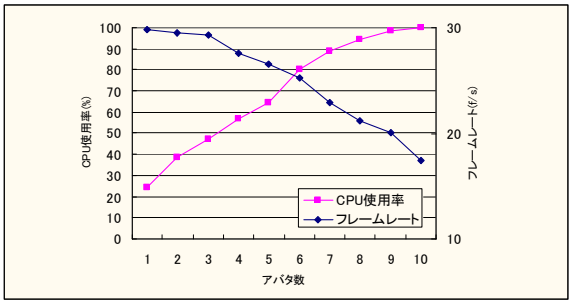
ネットワークは，100Mbps の Ethernet 上で行った．

(1) クライアントの描画速度

実験方法は，利用者がアバタを通じて，5 秒以内に一度発言すると仮定して，発言時に必ず表情変化を行わせた時に，アバタの数を増加させていった場合の計測を行った．

Flash ドキュメントでは、一度フレームレートを決定すると，それがドキュメント全体に適用されるため、

Flash アニメーションの再生開始前にこのレートを設定しておく必要がある．今回はビデオレート(30f/s)に固定して，フレームレートの変化を測定する．【図 11】にアバタ数に対する，平均フレームレート(f/s)と平均 CPU 使用率(%)の結果を示す．一部屋あたり数名のユーザが同時使用するときを想定すれば，通常の PC で十分なフレームレートを獲得できている事がわかる．



【図 11】アバタ数に対するフレームレート及び CPU 使用率

(2) ネットワークトラフィック量

ユーザがチャットルームにログインし，全角 50 文字のテキストによる会話をし，その時にアバタの表情が変化した場合，クライアントー同期サーバ間で送受信されるトラフィック量を測定した．

【表 1】に，結果を示す．通常のブロードバンド回線であれば，送受信による影響は，殆どないと思われる．

【表 1】 ネットワークトラフィックの測定結果

サーバに送信するクエリ	要求クエリサイズ	受信サイズ
チャットルームのダウンロード(初回時)	0.4kB	217kB
表情表出時の変形パラメータ	0.7kB	0.3kB
会話時のテキスト情報(全角50文字)	1.2kB	0.3kB
イベントチェック(20ms毎)	0.4kB	0.3kB

(3) 同期サーバのパフォーマンス

同期サーバに対して，全角 50 文字の入力テキスト情報と表情パラメータをインターバルなく 1000 回，通知するとして，同時利用者数を増加させていった場合のサーバの平均スループット[request/s]，平均レスポンスタイム[ms]，平均 CPU 使用率(%)を測定した．結果を【表 2】に示す．同時利用者数が増えた場合でも，同期サーバ側のパフォーマンスは十分であるといえる．

【表 2】 同期サーバのパフォーマンス測定結果

同時接続数	スループット[request/s]	レスポンスタイム[ms]	CPU使用率(%)
10	288.7	34.6	15.5
25	281.5	88.8	18.4
50	274.5	181.5	21.3
100	246.9	205.5	25.5
1000	208.1	279.3	32.6

以上をまとめると，クライアント側では，3D 顔画像合成による描画処理に影響して，フレームレートが下がる結果となったが，閉じたコミュニティでの利用であれば，アバタ数の上限を制限する事で対応は可能である．また，ネットワークトラフィック量，同期サーバの処理能力は十分と言える．

4. おわりに

本稿では，3D 顔画像合成によるコミュニケーションシステムについて述べた．システム的なパフォーマンスは，十分獲得できたと考える．

今後は，相手に効果的に感情が伝達できたかの定性評価を行うとともに．入力テキストと 3D 顔画像合成でのコミュニケーションだけではなく，音声によるコミュニケーションも付加して行く予定である．

[1] Macromedia Flash MX 2004: http://www.macromedia.com/jp/software/flash

[2] 佐藤，阿久津，南，外村 :「CyberCoaster: 映像の時空間直感的操作による可変速再生方法とその応用」，情処論 Vol.40, No.2, pp.529-536 (1999)

[3] Ekman, P. : Expression and the Nature of Emotion. In Scherer, K. and Ekman, P. (EDS.), Approaches to Emotion, pp. 319-343, Erlbaum, Hillsdale, NJ(1984).

[4] Nakanishi, H., Yoshida, C., Nishimura, T. and Ishida, T. : FreeWalk: A Three Dimensional Meeting Place for Communities, In Ishida, T. (Ed.), Community Comuting, pp. 55-89, John Wiley & Sons (1988)

[5] Ekman, P. and Friesen, V. W. :「Facial Action Coding System」, Consulting Psychologist Press, (1977)

[6] 宮下，森島他 :「仮想人物との対話を実現するための音声から画像への実時間メディア変換システムの研究」，信学技報 TECHNICAL REPORT OF IEICE. MVE95-62 (1996-03)