論　文

HyperMask：3次元顔モデルを用いた仮面の構築

四倉　達夫[†,††]　Kim Binsted[†††]　Frank Nielsen[††††]
Claudio Pinhanez[†††††]　鉄谷　信二[††]　中津　良平[††]
森島　繁生[†]

HyperMask: Reactive Talking Head for Storytelling

Tatsuo YOTSUKURA[†,††], Kim BINSTED[†††], Frank NIELSEN[††††],
Claudio PINHANEZ[†††††], Nobuji TETSUTANI[††], Ryohei NAKATSU[††],
and Shigeo MORISHIMA[†]

あらまし　HyperMask とは従来単一の顔表情や人物を表現する仮面の概念を進化させ，一つの仮面からあらゆる表情や人物を自由に生成及び表現可能なシステムである．本システムを用いることで，その仮面を装着した役者の表現の幅や新しい演出方法が生み出されていくと考えられる．顔の表出手法として，仮面に装着された五つの LED を，カメラにより追跡することで仮面の位置及び方向を求め，プロジェクタによって算出されたパラメータをもとに顔画像の投影を行う．また投影されている顔画像は演技者の音声を分析することによりリアルタイムで音声同期して口形状のアニメーションを行い，顔表情や人物の切換はユーザが任意に選択可能である．本論文ではHyperMask システムを用いた演出支援装置を紹介し，新たな仮面の表現技法の確立を目指す．

キーワード　仮想空間，顔合成，ニューラルネットワーク，ホログラフィ

1. まえがき

HyperMask は仮面を装着した演劇への支援手法であり，観客に対する演技者の表現力を大きく広げ，新たな演出が構築可能なシステムである．俳優や演劇者などに，白色の仮面を装着させ，プロジェクタによって表情・口形状変形及び人物の切換が可能な 3 次元顔モデルを投影しストーリーや場面，状況の変化に応じて容易に操作ができるように構成されている．また，演技者の動きに応じて仮面に正しく顔モデルを投影でき，自由度の高いシステムとなっている．

一般的に仮面と呼ばれているものは文化・宗教・地域によって様々なものが存在している．また仮面制作において写実的な表出を施してあるものもあれば逆に抽象的表現を施したものもあり多種多様である．それらは一様に単一の表情・人物が描かれており，演技するキャラクタのストーリー上での素性や立場に応じて使い分けている．また日本の古典芸能である能面のような演技者の動きや観客の視線方向，照明条件，ストーリーや音楽等様々な舞台環境から観客の心理状態，創造により仮面に内的な表情を付加させ演出を行う手法も存在する [1]．HyperMask の演出手法は従来の手法と異なり，外的な表情の変化，そして役柄自体の切換が可能で従来の仮面の概念を超えた自由度の高い演出が表現できると考えられる．また単純な表情から内的な感情を付加させるような複雑な表情を容易に表出可能にするため，仮面に投影する顔画像はリアルな顔モデルを用いるよう工夫した．それにより従来の仮面を使った独特な演出を熟知していない演技者でも簡単に表情の変化が操作可能で，観客もまた直感的に演技

† 成蹊大学工学部，武蔵野市
Faculty of Engineering, Seikei University, 3–3–1 Kichijoji-kitamachi, Musashino-shi, 180–8633 Japan
†† ATR 知能映像通信研究所，京都府
ATR Media Integration & Communication Laboratories, 2–2–2 Hikaridai, Seika-cho, Soraku-gun, Kyoto-fu, 619–0288 Japan
††† I–Chara，東京都
I–Chara K.K., 2–34–1 Uehara, Shibuya-ku, Tokyo, 151–0064 Japan
†††† ソニーコンピュータサイエンス研究所，東京都
Sony Computer Science Laboratories, Inc., 3–14–13 Higashigotanda, Shinagawa-ku, Tokyo, 141–0022 Japan
††††† IBM T.J. ワトソンリサーチセンタ，米国
IBM T.J. Watoson Research, 30 Saw Mill River Rd. (Route 9A) – Hawthorne, NY 10532, U.S.A.

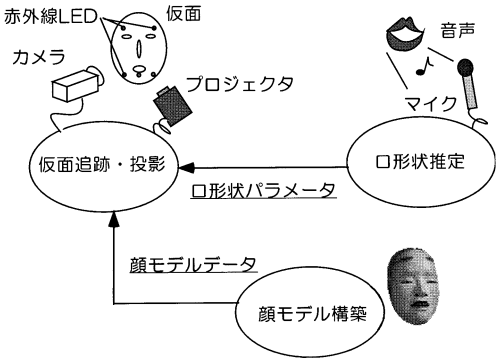


図 1　システム構成
Fig. 1　System feature of HyperMask.

者の表情を読み取ることができ，没入度の高い演劇空間が広がると考えられる．

本システムの構築にあたり，(1) 演技者が装着している仮面の動きを正しく追跡し，仮面上に顔画像を正しく投影する手法，(2) 仮面に投影する顔モデルの生成及び口形状・表情の制御法，(3) 音声から口形状への推定，計三つの基盤技術を用いている（図 1）．(1) において，本論文では簡単かつ短時間でキャリブレーションを行え，また精度の高い追跡・投影が可能な方法を提案し，(2) ではフレシキブルな顔モデルの構築を目指し，3 次元ワイヤフレーム上に顔のテクスチャ画像を容易にフィッティング可能なツールを開発した．また口形状・表情変化に必要なワイヤフレームの特徴点制御ルールを紹介する．(3) では仮面上の顔画像をリアルタイムにて制御を行うため，演技者の声をニューラルネットワークによって分析を行い，リアルタイムに音声と同期させ合成を行った．表情変形及び投影を行う顔モデルの変更は演技者がマニュアルで操作可能である．

紹介した三つの基盤技術はHyperMask を構築するために大変重要な要素となっているが，本システムのみ利用可能ではなく，他の分野へ容易に応用・転用可能できると期待される．例として技術 (1) を用いて “The Office of the Future” [2], [3] と呼ばれる実世界と仮想空間との共有インタラクティブスペースの構築に利用できると考えられる．また技術 (2), (3) では顔画像生成技術は制御パラメータのみで表情・口形状変形が可能であることから MPEG-4 [6] プロトコルに類似したテレビ電話やサイバスペース上での低ビットレートコミュニケーションシステムなど幅広い応用が期待できる．

以下，これらの基盤技術を **2.** (1) 仮面の追跡・投影，**3.** (2) 顔モデル構築，**4.** (3) 口形状推定，と順に追って紹介し，**5.** で実際にHyperMask システムを用いたプロトタイプの演出支援システムの構築，そしてプロトタイプを用いてデモンストレーションを行い，そのシステム評価結果を述べていく．

2.　仮面の追跡・投影

本章では 1 台のカメラとプロジェクタを用いた基盤技術の一つである (1) 仮面の追跡・投影について述べる．演技者は舞台上で静止していることはほとんどなく，演出に応じた動きを行う．そのために投影する顔モデルと仮面とが常に正しく投影され，かつ演技者の動きの制約を可能な限りなくしたシステム設計が求められる．またカメラとプロジェクタのキャリブレーションに関しても利用者の専門知識の必要なく短時間で各種パラメータが設定不要の単純な操作が望まれる．そこで本手法ではそれらの問題を解決すべく，次に述べる手法でキャリブレーションを行った．

2.1　キャリブレーション [4]

カメラ対と平面上に存在する観測点との関係を調べる際，一般的にホモグラフィ [7]（共線変換とも呼ばれる）が用いられている．ホモグラフィは実空間中の同一平面上に乗る複数点を 2 台のカメラで撮影したときの画像間での対応を表現し，ホモグラフィで記述される対応は同一平面上のみ有効で，2 台のカメラの位置や対象となる平面に依存する 3 行 3 列の行列で定義される．

投影モデルを考える際，基礎的な概念として理想状態のピンホールカメラモデルがよく知られているが，プロジェクタもまた理想状態でのピンホールモデルとして考えても差し支えない（図 2）．H をホモグラフィとするとプロジェクタ画像フレーム $\bar{p} = (x_p/w_p, y_p/w_p)$ とカメラ画像フレーム $\bar{c} = (x_c/w_c, y_c/w_c)$ との関係は次式のように示される．ただし，座標系は同次座標 $p = (x_p, y_p, w_p)$，$c = (x_c, y_c, w_c)$ を用いる．

$$p = \begin{pmatrix} x_p \\ y_p \\ w_p \end{pmatrix} = Hc = H \begin{pmatrix} x_c \\ y_c \\ w_c \end{pmatrix} \tag{1}$$

もし両画像平面上の 4 点の基準点となる座標がわかれば，ホモグラフィは完全に定義できる．本手法では仮面上に設定した 4 点とプロジェクタから仮面表面の 4

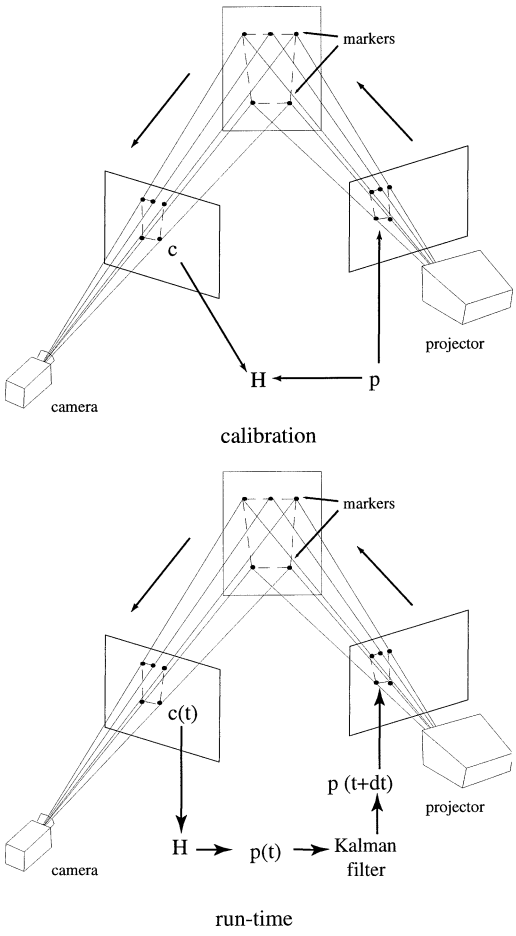


図 2　キャリブレーションと実行時のプロセス
Fig. 2　Calibration process and run-time process.

点の位置関係が対応した 4 点の画像をマニュアルで合わせることでカメラとプロジェクタ間のホモグラフィを求めた（図 2）.

プロジェクタの 4 点の同次座標（$w_p = 1$ とする）を

$$p_i = (x_p^i, y_p^i, 1) \qquad i = 1, 2, 3, 4 \tag{2}$$

とし，次にカメラの 4 点の同次座標（$w_c = 1$ とする）を

$$c_i = (x_c^i, y_c^i, 1) \qquad i = 1, 2, 3, 4 \tag{3}$$

とした.

ホモグラフィ行列 H を

$$H = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix} \tag{4}$$

とし，同次座標 (x, y, w) が (x'', y'', w'') に変換されると定義する．同次座標系は定数倍の自由度があり，ホモグラフィ行列も定数倍の自由度があるため $h_9 = 1$ としておくことができ，8 パラメータの変換として考えることが可能である.

$$x' = \frac{x''}{w''} = \frac{h_1 x + h_2 y + h_3}{h_7 x + h_8 y + 1} \tag{5}$$

$$y' = \frac{y''}{w''} = \frac{h_4 x + h_5 y + h_6}{h_7 x + h_8 y + 1} \tag{6}$$

上式は以下のようにも示される.

$$x' = h_1 x + h_2 y + h_3 - h_7 x x' - h_8 y x' \tag{7}$$

$$y' = h_4 x + h_5 y + h_6 - h_7 x y' - h_8 y y' \tag{8}$$

よって次式のような線形系で示され，H を 8 行 8 列 S の逆行列で求めることができる.

$$S = \begin{pmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1' x_1 & -x_1' y_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -y_1' x_1 & -y_1' x_1 \\ x_2 & y_2 & 1 & 0 & 0 & 0 & -x_2' x_2 & -x_2' y_2 \\ 0 & 0 & 0 & x_2 & y_2 & 1 & -y_2' x_2 & -y_2' x_2 \\ x_3 & y_3 & 1 & 0 & 0 & 0 & -x_3' x_3 & -x_3' y_3 \\ 0 & 0 & 0 & x_3 & y_3 & 1 & -y_3' x_3 & -y_3' x_3 \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -x_4' x_4 & -x_4' y_4 \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -y_4' x_4 & -y_4' x_4 \end{pmatrix}$$

$$H'^T = (\, h_1 \quad h_2 \quad h_3 \quad h_4 \quad h_5 \quad h_6 \quad h_7 \quad h_8 \,)$$

$$Z^T = (\, x_1' \quad y_1' \quad x_2' \quad y_2' \quad x_3' \quad y_3' \quad x_4' \quad y_4' \,)$$

$$SH' = Z \tag{9}$$

また

$$x_i' = x_p'^i,\ y_i' = y_p'^i,\ x_i = x_c^i,\ y_i = y_c^i,\ i \in \{1, 2, 3, 4\}$$

と置き換えることができる.

先ほども述べたとおり，H' を求めるとき S の逆行列が必要であるが，特異値分解や擬似逆行列を用いて逆行列を求める必要性もあり，安定した H' を求めることが困難な場合がある．他の手法として四つ以上の点を配置して最小 2 乗法や Kanatani [5] らの手法があるがこれ以上点を増やすことは投影する顔画像の印象に影響が出るため好ましくない．そこで本手法ではカメラ，プロジェクタ両方とマスクの距離が十分に遠いと仮定することにより，アフィン変換によって H を導出する.

4 点のカメラ，プロジェクタ

$$P = (p_1^T, p_2^T, p_3^T, p_4^T) \tag{10}$$

$$C = (c_1^T, c_2^T, c_3^T, c_4^T) \tag{11}$$

とし，$P = HC$ を用いて H を求めると

$$H = PC^T(CC^T)^{-1} \tag{12}$$

のようになる．実行時はカメラ画像 $c = (x_c, y_c, 1)$ と，キャリブレーションの際に求まった H を用いて p を求めることができる．本手法で用いたキャリブレーションは予備実験の結果 4 点のみを用いて非常に安定した処理を実現しており，キャリブレーションに要する時間も短時間で済む．またカメラやプロジェクタ固有のパラメータを必要としない．

2.2 投影面の追跡

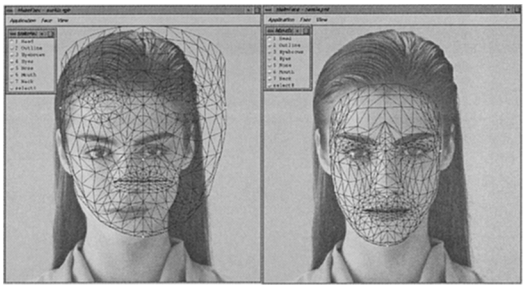
投影面の追跡を行う際，本論文では赤外線 LED を用い，カメラ部には赤外線フィルタを装着させている．また，4 点 LED のラベリングを行うためにもう 1 点ラベリング用の LED を用意し，計五つの LED を仮面に付けた．投影面の追跡及びトラッキングは非常に安定していると前節で述べたが，マスクの移動が非常に早い場合，投影画像に「遅延」が生じてしまう問題がある．カメラ画像の取込み時間を t とし，プロジェクタへの投影 $t + dt$ とすると dt の遅延が生じてしまう．この問題を解決するため本論文ではカルマンフィルタ [8] を用いた（図 2）．パラメータ dt はカメラ画像の取込みからプロジェクタの投影までの平均時間とし，経験的に求めた．カルマンフィルタを用いた場合と用いない場合を実験で比べてみたところ，フィルタによって遅延現状が軽減されており満足のいく結果となった．

3. 顔モデル構築

HyperMask では実際の人物の顔と同レベルに近いクオリティでの表出を目的にしているため，微妙な表情が表出可能で観客に対し違和感を与えない顔モデル，また口形状や表情の制御ルールが必要となる．本章では基盤技術（2）にあたる表示用顔モデルの作成法の紹介，また作成した顔モデルの表情や口形状の制御方法を紹介する．

3.1 3 次元顔モデル

リアルな顔モデルの製作のため，演技に使用する対象人物の正面画像を三角形ポリゴンで構成させる顔の標準ワイヤフレームモデルをマニュアル整合し，個人モデルを作成する．このモデルは約 850 ポリゴンの三角形パッチにより構成されていて，格子点数は約 480 点から形成される．ポリゴン数は形状の変化の際，演算量及びレンダリングの処理時間に直接関係する．ここでは実時間でのアニメーション実現のため，動きの変化の激しい部分（唇，眉，眼周辺部）にのみ細かいポリゴンを割り当て，全体的な演算量の軽減を行っている．そしてこのモデルにテクスチャマッピングを施すことによって顔合成画像を作成する．また唇を開けたときを考慮し歯及び口内部のモデルを追加した．歯のモデルは白色系のグローシェーディングを施しており，口内部は袋状のモデルにペイントツールで作成したテクスチャイメージを付けた．



(a) Initial Model (b) Fitted Model

図 3 整合ツールウィンドウ
Fig. 3 Fitting tool's window.

顔モデルを対象人物に整合した様子を図 3 に示す．顔モデルの整合を容易に行うため，GUI ツールを開発した [9]．まず演出の際に必要な顔画像を読み込む．顔モデルのワイヤフレームモデルの格子点を動かし画像と顔モデルの整合を行う．点の移動ははじめマクロに制御して，次第に細かく位置合せでできるように考慮されている．また実際に表情変形してみて，不自然な部分はインタラクティブに位置修正できるように配慮されている．特に目と唇の部分は表情変形に重要であるため綿密な整合が必要である．図 3 (a) は整合前の編集画面であり，(b) は整合された後の画面を示している．このツールを用いて顔モデルを完成させる所要時間は全くの初心者でも約 5 分程度で完成できる．

3.2 表情及び口形状のパラメータ化

仮面に投影された顔モデルの表情や口形状変化を表現する顔画像を構築するために，3 次元顔モデルの幾何学的変形の基準となる特徴点の設定と，その移動量の記述，そして特徴点の周囲の格子点の移動規則などを定める必要がある．ここではモデル変形の基礎とな

る表情と口形状の制御パラメータについて述べる．

3. 2. 1 表情パラメータ

表情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）[10] と呼ばれる動きの方向を解剖学的に考慮して顔の表情を AU（Action Unit）と呼ばれる 44 個の基本動作に分類している．あらゆる表情は AU の組合せで表現できるとされ，FACS は表情記述単位として顔画像の分析，合成分野で広く用いられている．各 AU は顔面上の特徴点の 3 次元移動ベクトルとして定義されている．表情変化は 3 次元モデルの特徴点の AU の強さによって移動させ，特徴点以外の格子点は，特徴点の移動に基づく補間によって制御される．特徴点は 48 点設定して，各特徴点の位置は唇・眉・目輪郭部と頬周辺部，額上部と耳に配置している．感情の種類としてこの AU の組合せによって表現された，怒り，喜び，悲しみ，嫌悪，驚き，おそれの 6 基本感情を標準として用意した．もちろん，この AU のポリゴン数は形状の変化の際，演算量及びレンダリングの処理時間に直接関係する．ここでは実時間でのアニメーション実現のため動きの変化の激しい部分にのみ細かいポリゴンを割り当て，全体的な演算量の軽減を行っている．このモデルにテクスチャマッピングを施すことによって顔合成画像を作成する．また，歯及び口内部のモデルを追加した．編集によってユーザ自身で感情をカスタマイズするための AU エディタも用意されている．図 4 に基本 6 感情の合成画像の一例を示す．これはあくまで標準として用意するもので，ユーザによるカスタマイズは容易に実行可能である．

3. 2. 2 口形パラメータ

発話時の口形状を表現するために，先に述べた AU とは異なる，口領域の変形に限定したパラメータを用いる．パラメータは五つの母音（/a/，/i/，/u/，/e/，/o/）と閉口の口形状を基準とし，すべての口形状はこれらの補間によって再現できると仮定している．

口領域の動きを少数のパラメータで表現するために，口領域の制御点のパラメータとして図 5 のような 13 個を定めた．3 次元計測結果に基づいて，この制御点自体の移動量の算出，更に制御点以外の格子点の移動量算出ルールを定めた．この 13 個の座標値によって，唇の形状を一意に決定することができる．図 6 にこの口形パラメータによって表現された口形/u/の合成画像を示す．

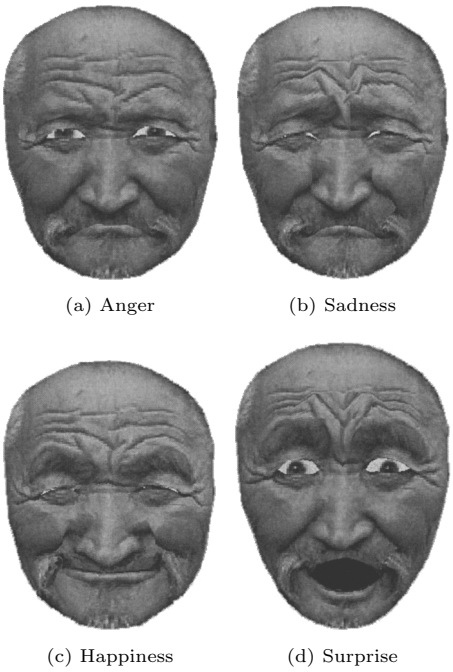
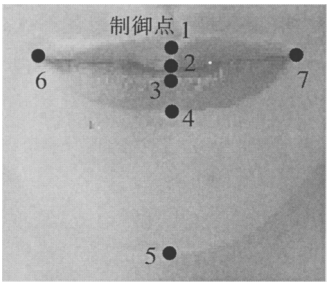


図 4 顔モデルによる各表情合成画像
Fig. 4 Example of face synthesis images.



制御点	口形パラメータ	制御
1	1	上唇上側の縦方向の動き
	2	上唇上側の奥行方向の動き
2	3	上唇下側の縦方向の動き
	4	上唇下側の奥行方向の動き
3	5	下唇上側の縦方向の動き
	6	下唇上側の奥行方向の動き
4	7	下唇下側の縦方向の動き
	8	下唇下側の奥行方向の動き
5	9	あごの縦方向の動き
6	10	口角の縦方向の動き
	11	口角の横方向の動き
	12	口角の奥行方向の動き
7	13	唇の横方向の開き具合

図 5 口形パラメータ
Fig. 5 Location of mouth parameters.

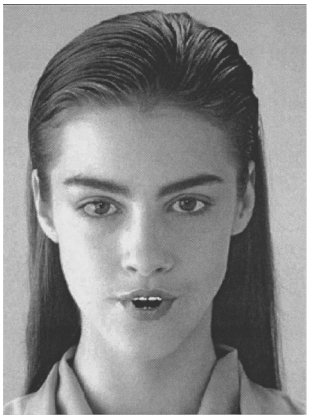
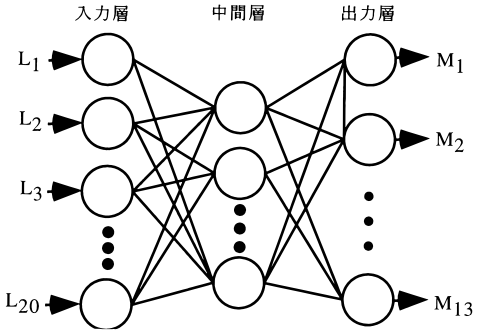


図 6　口形/u/の合成画像
Fig. 6　Face synthesis image of vowel ‘u.’



L_n: LPC cepstrum, M_n: Mouth parameter

図 7　口形パラメータへの変換に用いるニューラルネットワーク
Fig. 7　Neural Network for conversion to mouth shape parameters.

4.　音声情報から口形状の推定 [11]

三つ目の基盤技術となる口形状の推定手法は顔モデルの口形状をリアルタイムに決定するため，ユーザから入力された音声をフレームごとに分析することによって，毎フレーム口形パラメータを推定する．特徴パラメータとして計算時間が比較的少なく，また発話者の声動特性と放射特性の特徴を表現していると考えられる LPC ケプストラム係数とした．入力音声は 16 [kHz]，16 [bit] とし，分析フレーム長及び周期は 32 [ms] で切り出す．

LPC ケプストラムから口形パラメータへの変換は図 7 のような 3 層フィードフォワード型ニューラルネットワークを用いている．入力層は LPC ケプストラム次数と同じ 20 ユニット，出力層は 13 個の口形パラメータに相当する．更に中間層は経験的に 20 ユニットとした．学習パターンは 5 母音の LPC ケプストラムとそれぞれの発話時の口形パラメータ，及び無発音時の周囲の環境雑音から求めた LPC ケプストラム係数と閉口口形とした．収束までに 100 万回の学習を行った．このニューラルネットの重み係数は基本的に話者依存性が強く，話者ごとに事前に学習を行う必要がある．この問題を解決するために後述する話者適応処理によってこの学習を省略することもできる．

4.1　話 者 適 応

本システムは不特定多数の方が利用すると考えられるが，ユーザが更新されるたびにニューラルネットによって学習を行うことは非能率的である．そこであらかじめ収録した 100 人分の学習データで重み係数のデータベースを構築した．この中からユーザに最適な重み係数を自動的に選択する．新しいユーザには，実験開始直前に 5 母音を発生してもらい，データすべての中から一つずつ選択された重み係数によって順次口形状推定を行い，基準の 5 母音の口形状に最も近いものを生成できる重み係数をその人物の最適な係数と判断して，話者適応を実施した．

4.2　口形状推定評価

1995 年 8 月にロサンゼルスで行われた ACM の SIGGRAPH ’95 において，インタラクティブデモ展示を行った [12]．このデモでは，会場に訪れた人物の顔正面画像と 5 母音の音声をその場で取り込み，モデル整合と話者適応処理の後に，リアルタイムでマイクから入力された音声を分析して，口形を合成する処理を行い，合成された顔画像を通じて 2 者間で対話を行うというものであった．このデモにおいて，来場者 160 人の整合処理と話者適応を実施し，すべての人物において自然な口形状と表情の合成が可能であることが明らかとなった．なお，この際の表情合成速度は毎秒 10 フレームであり，すべての外国人を対象として対話は英語で行われた．整合処理は経験のある人物によって実施されたが，平均 1 分程度の所要時間であった．

5.　プロトタイプ

前章まで述べた手法を用いて図 8 のようなプロトタイプを構築した．プロトタイプを構築するにあたり，演技者が舞台内を自由に動け，演技に支障をきたさないシステムを構築するために図 9 のようなカート内にカメラやプロジェクタ，コンピュータ等を収納し，装

図 8　プロトタイプ
Fig. 8　Prototype of HyperMask.

#1: Camera (covering Infrared Filter)
#2 Key-pad #3 Projector.

図 9　カート内部
Fig. 9　Trolley inside.

置自体が自由に移動可能なポータブルシステムを構築した．

5.1　システム構成

処理用のワークステーションとして，SGI 社製 Indigo2（MIPS 10000，123 MByte，IRIX6.5），を使用した．このワークステーション 1 台で（1）仮面の追跡・投影，(3）口形状推定のプロセスを並列処理している．仮面追跡用カメラ（Sony EVI-G20），顔画像投影用プロジェクタ（Sony），そして赤外線 LED が埋め込まれた白色の仮面を用いた．仮面の目にあたる部分は演技者の視界を確保するため数箇所小さな穴を空けた．これらの穴は直径 1.5 mm 程度であるため投影画像に影響を受けず，プロジェクタの輝度による演技者への影響は今回使用したプロジェクタの輝度が 50ANSI ルーメンで演技者の目に与える負担は非常に少なく，また予備実験から視界に影響を与えないことがわかった．そして演技者が台詞を発するとき声がこもり，観客に聞こえないおそれがあるため，仮面内側には小型マイクを付け，カート内にあるスピーカと接続させ声量の確保を行っている．

プロトタイプを運用する前にいくつかの準備を行う必要がある．まず基盤技術（2）を用いて演出に用いる顔モデルの制作を行う．次に口形状の推定を行うために必要なニューラルネットの重みデータの学習を行う．声の収録の際，演技者には仮面を装着してもらいデモ環境と同じ状態で録音した．今回演技者が 2 名であり，最適な推定を行うため先述した話者適応の手法は使用しなかった．最後にカメラのキャリブレーションを行って準備が完了する．

演技者はカートを押しながら舞台を移動しパフォーマンスを行い，また観客とのインタラクションを行う．仮面に投影された顔は様々なストーリー展開や口調，観客とのインタラクションによって変化する．カート上のカメラは常時演技者の仮面を追跡し，プロジェクタもまたリアルタイムに顔表情を合成させたモデルを投影する．演技者が発話したセリフすべて最適な口形状へとリアルタイムに処理され，顔モデルの口形が生成される．顔表情や投影する人物の顔画像の変更はカート上に装備してあるテンキーによって演技者が任意に変更が可能である．

5.2　評価実験

本プロトタイプを用いて，1999 年 8 月 SIGGRAPH '99 のエマージングテクノロジーにて実際に一つのオリジナルストーリーを作成し，観客とのインタラクティブなコミュニケーションも取り入れたデモンストレーションを行った [13]．顔画像は図 10 のように投影され，ストーリーやアドリブによって口形状をリアルタイムに推定，合成を行い，表情表出及び演出する役の切換はカートに備え付けているキーパッド（図 9）によって任意に変更可能となっている．図 11 にデモンストレーションの様子を示す．

1 回のパフォーマンスで約 10 人から 30 人の観客が訪れ，開催期間中約 1000 人参加した．デモに参加した観客の大多数が本システムの演出法に対し大変興味をもち，斬新かつ応用性をもつシステムであると好評を得た．また投影した顔表情の合成フレームレートは毎秒 8〜10 フレーム前後ではあったが仮面に投影された顔モデルの口形状に対して同期や表情表出が自然に

図 10　顔画像投影例
Fig. 10　Projected face on mask.

図 11　デモンストレーション風景
Fig. 11　Snapshot of demonstration.

表現されているとの意見を頂いた．実際に音声と口形状との遅延時間を計測したところ約 30 ms ほど音声の方が早く出力していたが同期には影響しなかった．

6. むすび

本論文ではプロジェクタによって口形状や表情が変化し，投影する人物の顔が選択可能な仮面を用いた演技支援システム “HyperMask” を提案した．本システムは（1）仮面追跡,（2）顔モデル構築,（3）口形状推定，以上三つのプロセスで構成され，各々のプロセスで用いた手法の考察は以下のとおりである．

（1）　仮面追跡・投影

赤外線 LED を装着した白色の仮面のトラッキング及びプロジェクタによる投影に関し，ホモグラフィを用いることで 4 点（LED のラベル付けを含めると 5 点）の LED のみで容易かつ短時間にキャリブレーションでき，仮面の追跡の精度も良い結果がでた．仮面 LED のキャプチャから投影までのプロセス処理時間で発生する「遅延」の問題はカルマンフィルタを用いることで，ほぼ仮面の動きと顔画像とが遅延なく投影されていることが確認できた．このアルゴリズムでの仮面の回転運動の許容範囲は実験結果から，ロール ϕ・ピッチ θ・ヨー ψ で示すと $0^\circ \leq \phi < 360^\circ$, $-20^\circ \leq \theta \leq 20^\circ$，$-30^\circ \leq \psi \leq 30^\circ$ である．すべての LED がカメラによってとらえなくてはならないためにこのような制限があるが，演技者の動作に支障が生じることはなかった．

（2）　顔モデル構築

HyperMask で使用する仮面の顔画像はリアルな顔を再現することをが目的となっている．そこで標準ワイヤフレームモデルを用意し演技用の顔画像とのフィッティングを行い，ワイヤフレームモデルにあらかじめ設定した表情・口形状制御ルールで表出を行った．これらのルールは顔の一つひとつの基本的な動きをパラメータとして定義したので基本 6 感情のみならず複雑な表情やすべて音素発音時の口形状に対応できる．ワイヤフレームモデルと顔画像とのフィッティングでは GUI ベースの整合ツールを用意し簡単に処理可能である．

（3）　口形状推定

演技者の音声情報からニューラルネットワークを用いることで母音推定が可能となった．母音のみの推定ではあるが先に述べたデモンストレーションで英語発生時の口形状評価を行ったところ，自然な表出が出ているとの回答を頂いた．また推定に要する処理時間も大変少ないため音声と合成画像との同期も問題がなかった．今後母音のみならず唇の形状と密接な関係をもつ他の音素（例えば破裂音の/Ba/，/Pa/）の推定を行いより精度の高い推定システムの構築を考えている．

そしてこれらを用いて実際にHyperMask システムのプロトタイプを製作し，演技者が舞台内を移動できるようカート状のポータブルなシステムを提案した．観客とのインタラクションやオリジナルストーリーのデモンストレーションを行うことでHyperMask システムの実現性が確認でき，従来までの仮面の概念を超えた全く新しい演出技法の有効性が明らかとなった．今後の展開として，システム全体の軽量・小型化，使いやすさの向上を図っていく．現状ではワークステーションやプロジェクタ等をカートに入れているが，近年の急速な電子機器の発展により実現は困難ではない．

また表情の操作，人物の入換にはキーパッドを使っているが今後，操作デバイスの改良や音声による感情推定システムによる自動化を考慮していく．

本システムはカメラによる追跡・投影，顔合成，音声分析技術を融合している．そのためこれらの技術はHyperMaskシステムのみならず今後，様々な分野への応用化が可能であると考えられる[14]．現状では演劇に特化したシステム構成となっているが，カメラ追跡・投影技術（追跡プロセス）を用いることで先に述べた“The Office of the Future”への転用も可能である．また顔合成・音声分析システム（音声分析・投影プロセス）を用いてフェース・トゥ・フェースでの多人数コミュニケーションシステム，電子会議システム等への応用化も検討中である．

文　献

[1] M.J. Lyons, A. Plante, M. Kamachi, S. Akamatsu, R. Campbell, and M. Coleman, “Viewpoint dependent facial expression recognition: Japanese noh masks and the human face,” Proc. 22nd Annual Conference of the Cognitive Science Society, pp.322–327, 2000.

[2] R. Rasker, G. Welch, M. Cutts, A. Lake, L. Stesin, and H, Fuchs, “The office of the future: A unified approach to image-based modeling and spatially immersive displays,” SIGGRAPH Annual Conference Proceedings, pp.179–188, 1998.

[3] C. Cruz-Neira, S.J. Daniel, and T.A. DeFanti, “Surround-screen projection-based virtual reality: The design and implementation of the CAVE,” Computer Graphics, SIGGRAPH Annual Conference Proceedings, pp.135–142, 1993.

[4] C. Pinhanez, F. Nielsen, and K. Binsted, “Projecting computer graphics on moving surfaces: A simple calibration and tracking method,” SIGGRAPH Sketches and Application, p.266, 1993.

[5] K. Kanatani, “Optimal homography computation with a reliability measure,” Proc. MVA ’98, IAPR Workshop on Machine Vision Applications, pp.426–429, 1998.

[6] E. Patajan, “Approaches to visual speech processing base on the MPEG-4 face animation standard,” Proc. ICME2000, 2000.

[7] O. Faugeras, Three-Dimensional Computer Vision: A Geometric Viewpoint, The MIT Press, Cambridge, Massachusetts, 1993.

[8] A. Gelb, Applied Optimal Estimation, The MIT Press, Cambridge, Massachusetts, 1974.

[9] 森島繁生，八木康史，金子正秀，原島　博，谷内田正彦，原　文雄，橋本周司，“顔の認識・合成のための標準ソフトウェアの開発,” 信学技報，PRMU97-282, 1998.

[10] P. Ekman and W.V. Friesen, Facial Action Coding System, Consulting Psychologists Press, 1978.

[11] S. Morishima, “Modeling of facial expression and emotion for human communication system,” Displays, vol.17, pp.15–25, 1996.

[12] S. Morishima, “Better face communication,” SIGGRAPH Sketches and Application, p.117, 1995.

[13] K. Binsted, “Virtual reactive face for storytelling,” SIGGRAPH Sketches and Application, p.186, 1999.

[14] K. Binsted, T. Misawa, S. Morishima, and F. Nielsen, “Denger hamster 2000,” SIGGRAPH Sketches and Application, p.81, 2000.

（平成 13 年 2 月 5 日受付，6 月 12 日再受付）

四倉　達夫　（学生員）

平 10 成蹊大・工卒．平 12 同大大学院修士課程了．現在同大学院博士課程在学中，及び（株）ATR 知能映像通信研究所研修研究員．超高精細顔モデルの構築・仮想空間上でのコミュニケーションシステムに関する研究に従事．平 12 本会学術奨励賞受賞．

Kim Binsted

is CEO of I-Chara Inc., a Tokyo-based mobile agent company (www.i-chara.com). Formerly, she was a researcher at the Sony Computer Science Laboratories, working on Human Computer Interaction and Artificial Intelligence (AI). She received her PhD in AI at the University of Edinburgh, and her BSc in Physics at McGill University, Montreal.

Frank Nielsen

received the B.S. and M.S. degrees from École Normale Supérieure (ENS) of Lyon in 1992 and 1994, respectively. He defended his Ph. D. thesis on “Adaptive Computational Geometry” prepared at INRIA Sophia-Antipolis under the supervision of Pr. Boissonnat in 1996. As a civil servant of the University of Nice (France), he gave lectures at the engineering schools ESSI and ISIA (École des Mines). In 1997, he served army as a scientific member in the computer science laboratory of École Polytechnique (LIX). In 1998, he joined Sony Computer Science Laboratories, Tokyo (Japan) as an associate researcher. His current research interests include computational geometry, algorithmic vision, combinatorial optimization for geometric scenes and compression.

Claudio Pinhanez

is a computer scientist and a media artist. He has been a researcher at IBM TJ Watson Research Center since 1999, and currently is part of the Pervasive Computing Group, working in the design and development of interactive spaces and on physical interfaces to information. Claudio got his PhD. from the MIT Media Laboratory in 1999 with Prof. Aaron Bobick, working on the design and construction of physically interactive environments. In particular, he investigated new paradigms for computational representation of human action and the problem of scripting stories in interactive environments. During his PhD. he created and produced innovative theatrical experiences involving computers interacting with human actors on stage, including the computer theater plays "SingSong" and "It/I." Claudio was a visiting researcher at ATR-MIC laboratory (Kyoto, Japan) in 1996 and at Sony Computer Science Laboratory (Tokyo) in 1998.

鉄谷　信二　（正員）

昭 55 北大工学部大学院修士課程了．同年電電公社（現 NTT）入社以来，ファクシミリにおける画像信号処理，電子写真記録，立体表示技術等の研究実用化に従事．平 3 ATR 通信システム研究所に出向，臨場感表示技術に従事．平 6 NTT に復帰，高速ネットワーク用アプリケーション開発に従事．平 12 ATR 知能映像通信研究所に出向，コミュニケーション環境生成に関する研究に従事．現在，同研究所第 1 研究室長．工博．

中津　良平　（正員）

昭 44 京大・工・電子卒，昭 46 同大大学院修士課程了．同年日本電信電話公社（現 NTT）武蔵野電気通信研究所入所．昭 55 横須賀電気通信研究所．主として音声認識の基礎研究，応用研究に従事．平 2 NTT 基礎研究所研究企画部長，平 3 NTT 基礎研究所情報科学研究部長．平 6 より ATR に移り，現在（株）ATR 知能映像通信研究所代表取締役社長．マルチメディア要素技術の研究及びマルチメディア技術を応用した通信方式の研究などに従事．工博（京大）．昭 53 年度本会学術奨励賞，平 8 IEEE Multimedia Systems and Computing '96 最優秀論文賞，平 9 ロレアル賞，平 11 映像情報メディア学会論文賞，平 11・12 テレコムシステム技術賞，平 11・12 日本バーチャルリアリティ学会論文賞，平 12 人工知能学会論文賞，平 13 文部科学大臣賞各受賞．平 13 IEEE フェロー．日本音響学会，情報処理学会，人工知能学会，画像電子学会，日本バーチャルリアリティ学会，映像情報メディア学会，日本芸術科学会，日本情報考古学会各会員．

森島　繁生　（正員）

昭 57 東大・工卒．昭 59 同大大学院修士課程，昭 63 同大学院博士課程了．工博．平 13 成蹊大学工学部教授，現在に至る．平 6 から 1 年間，トロント大学客員研究員．平 8 から通信放送機構 3 次元空間共有プロジェクトサブリーダー．明治大学非常勤講師．（株）ATR 音声言語通信研究所客員研究員を併任．本会論文誌編集委員．グラフィックス，ビジョン，マルチモーダルインタフェース等の研究に従事．平 4 本会業績賞受賞．