HYPER MASK: 3 次元顔モデルを用いた仮面の表現技法

四倉 達夫[1][2]　Kim BINSTED[4]　Frank NIELSEN[3]
Claudio PINHANEZ[5]　鉄谷 信二[1]　森島 繁生[1][2]

[1] ATR 知能映像通信研究所　[2]成蹊大学工学部
[3]ソニーコンピュータサイエンス研究所　[4] i-chara　[5] IBM T.J. Watson Research

HYPER MASK: Projecting Talking Head on Moving Surfaces

Tatsuo YOTSUKURA[1][2], Kim BINSTED[4], Frank NIELSEN[3],
Claudio PINHANEZ[5], Sinji TETSUTANI[1] and Shigeo MORISHIMA[1][2]

[1] ATR Media Integration & Communications Research Laboratories
[2] Faculty of Engineering, Seikei University　[3] Sony Computer Science Laboratories
[4] I-chara Inc.　[5] IBM T.J. Watson Research

概要

Hyper Mask とは従来単一の顔を表現する仮面の概念を進化させ、1 つの仮面からあらゆる表情や人物を自由に生成可能なシステムである。この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。白色に塗装された仮面上に 5 つの LED が装着されており、プロジェクタによって任意の顔画像を投影し、ストーリーや観客とのインタラクションによって仮面の顔を変化する。仮面に顔画像を正しく投影させるため、カメラから 5 つの LED を検出し仮面の位置を求めている。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に音声同期して口形状のアニメーションを行い、顔表情もまたユーザが任意に変形可能である。本稿では Hyper Mask 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を確立した。

１．はじめに

Hyper Mask は演劇用のデモンストレーション手法であり、演技者の観客に対する表現の幅を大きく広げ、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、その仮にプロジェクタによって表情・口形状変形および人物の切り替えが可能な顔モデルを投影する。

Hyper Mask で用いるリアルタイムで仮面などの物体へ正しく投影を行う技術は、他の分野へ容易に応用可能である。例として”The Office of the Future”[1][2]が挙げられる。動きのある不規則な形状の物体に動的に映像や情報を投影が可能な本システムを用いることで、実世界と仮想空間とが共有したインタラクティブなプレイグラウンドが構築可能である。

本システムは 2 つの基盤技術で構成されている。まず仮面上に顔画像を投影する手法、そして仮面に投影する顔モデルの表現方法である。前者では仮面上に 5 点の赤外線 LED を装備し、これらの点をカメラを用いて追跡を行い仮面のトラッキングを行う。後者ではフレシキブルな顔モデルを構築するため、3 次元ワイヤフレーム上に顔のテクスチャ画像を容易かつ短時間でフィッティング可能なツールを開発した。そして口形状・表情変化時にはワイヤフレーム

の特徴点を移動させプロジェクタによって投影を行う。また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う。表情変形およ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。

本稿では仮面のトラッキング手法、顔モデルの構築手法および Hyper Mask システムを用いたプロトタイプの演出支援システムを紹介し、実際にこの装置を使ってデモンストレーションを行った。そして本システムの評価および今後のシステムの方向性、展望を述べていく。

2．プロトタイプシステム

Hyper Mask のプロトタイプとして演技者が舞台内を自由に動けるよう、図 2-1 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(図 2-2)装置自体が自由に移動可能なポータプルシステムを構築した。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、また観客とのチャットを行う。仮面に投影された顔はさまざまなストーリーやインタラクションの内容、口調によって変化する。カート上のカメラは常時演技者の仮面を追跡し、LCD 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全ては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。本システム構成を図 2-3 に示す。

図 2-1. プロトタイプシステム

図 2-2. カート内部
#1. カメラ(赤外線フィルタ装備)
#2. プロジェクタ　#3. キーパッド

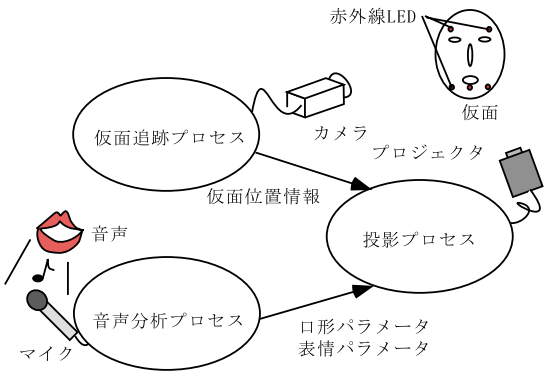


図 2-3. システム構成

3．キャリブレーション

2 つの異なるカメラから平面上に存在する点同士の関係はホモグラフィ[3]を用いることが知られている。ホモグラフィは投影空間内で 3 行 3 列の行列で定義される。

幾何学的観点から基本観測は ” 理想的 ” なピンホールプロジェクタやカメラらが同一である(図 3-1)。H はホモグラフィを示し、カメラ画像フレームとプロジェクタの画像フレームとが関係している。この意味はカメラ画像上の 2 次元の点

$$\bar{c} = (x_c / z_c , y_c / z_c)$$

は 2 次元の点

$$\overline{p} = (x_p / z_p, y_p / z_p)$$

と合わせると、プロジェクタ画像は以下のとおりに示される。

$$p = \begin{pmatrix} x_p \\ y_p \\ x_p \end{pmatrix} = Hc = H \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix}$$

もし両画像面上に存在する 4 点の 3 次元座標の投影法がわかれば、ホモグラフィは完全に定義される。カメラとプロジェクタ間のホモグラフィを明らかにするため、本研究ではマニュアルによってプロジェクタで投影された面上の画像と実際の表面とをマニュアルで合わせて必要な 4 点を単純に得る必要がある(図 3-1)。

これらの点は可視でき、実面と投影面を合わせるため、実面を動かすための方法があることを確認する必要があるが、投影された 4 点の同次の座標

$$p_i = \left(x_p^i, y_p^i, 1\right) \qquad i = 1,2,3,4$$

は任意に決定される。そこで本研究ではカメラ画像上の 4 点の座標をトラッキングシステムとして検討した。

$$c_i = \left(x_p^i, y_p^i, 1\right) \qquad i = 1,2,3,4$$

下記に示す 4 点の 2 式は一致する。

$$P = \left(p_1^T, p_2^T, p_3^T, p_4^T\right)$$

$$C = \left(c_1^T, c_2^T, c_3^T, c_4^T\right)$$

P=HC であるので解は

$$\overline{p} = \left(x_p / z_p, y_p / z_p\right)$$

である。実行中、簡単にカメラ画像内の点 *c=(xc,yc,1)* が求められ、ホモグラフィ *H* によって *p=Hc* が得られ、プロジェクタの画像表面の位置を計算可能である。驚くことに、このキャリブレーション工程はたった 4 点のみで安定した分析が可能である。また本研究ではカメラやプロジェクタの中心がほぼ一致しているため安定性は向上している。カメラやプロジェクタの固有なパラメータの計測は必要としない。

本研究で、映写面上に赤外線 LED をマーカとして採用した。このマーカは赤外線フィルタを装着したカメラによって容易にトラッキング可能である。

また安定した仮面への顔画像の投影を行うためにカルマンフィルタ[4]を用い、デモンストレーションの際、大変効果的な結果を示し、安定した投影が可能となった。

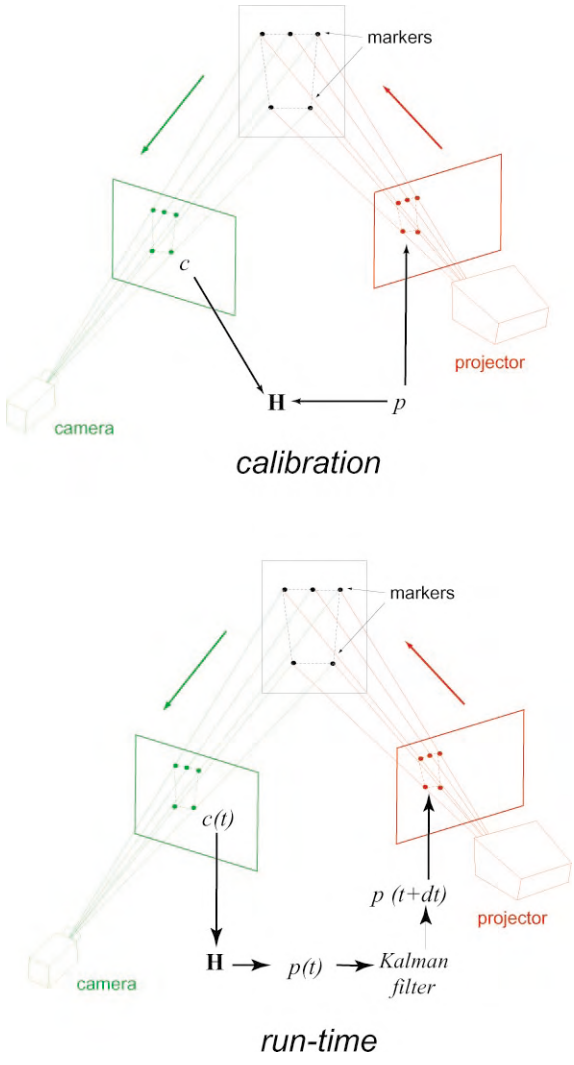
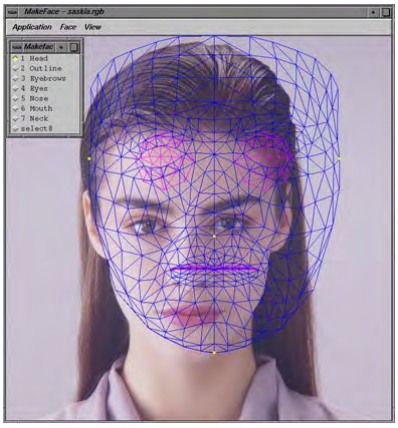


図 3-1. キャリブレーションと実行時のプロセス

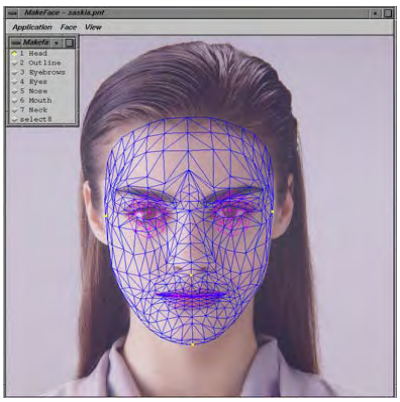
４．3 次元顔モデル

表情の表出と音声に同期した唇の動きを実現可能な顔画像をプロジェクタによって仮面へ投影するため、カメラから獲得した対象人物の正面画像に、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。

ホリゴン数は形状の変化の際、演算量およ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。



a) 整合前



b) 整合後

図 4-1. 整合ツールウインドウ

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、歯および口内部のモデルを追加した。

顔モデルを対象人物に整合した様子を図 4-1 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[5]。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。点の移動ははじめはマクロに制御して、次第に細かく位置あわ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 4-1. 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まったくの初心者でも約 5 分程度で完成できる。

５．表情およ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変化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ため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変形パラメータについて述べる。

５．１． 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)[6]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 の組み合わ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。感情の種類としてこの AU の組み合わせによって表現された、怒り、喜び、悲しみ、嫌悪、驚き、恐れの 6 基本感情を標準として用意した。もちろん、この AU のホリゴン数は形状の変化の際、演算量および

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、歯およ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 5-1 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。

５．２．口形状パラメータ

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。日本語の発音の口形状には、異なる発音でも同じような口形状となる同口形異音が多く存在する。よってすべての音韻に対応する口形を独立に用意する必要はない。また音声分析性能の限界から、細かい子音に関しての識別は困難である。特に日本語では、大半が母音区間と考えられるので母音区間の口形再現が自然さに大きく寄与すると考えた。

a) 喜び　　b) 驚き

図 5-1. 顔モデルによる各表情合成画像

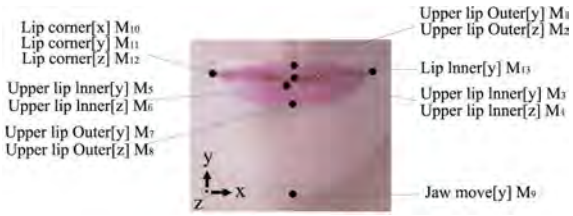


図 5-2. 各パラメータの位置

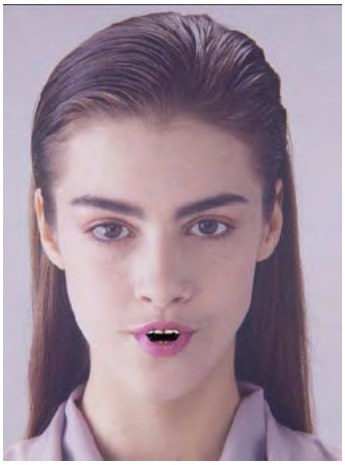


図 5-3. 口形/u/の合成画像

そこで 5 つの母音(/a/,/i/,/u/,/e/,/o/)と閉口の口形を基準とし、すべての口形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、口領域の制御点として図 5-2 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、さら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 5-3 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。

６．音声情報から口形の抽出[7]

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、たま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、分析フレーム長および周期は 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 6-1 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。さら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、

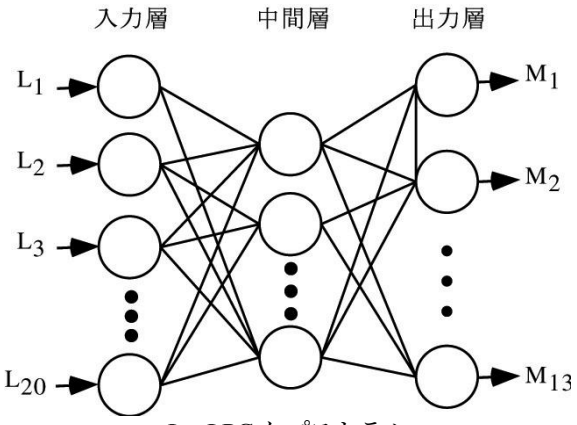


図 6-1. 口形パラメータへの変換に用いるニューラルネットワーク

およ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。

６．１．話者適応

被験者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 1 つずつ選択された重み係数によって順次口形推定を行い、基準の 5 母音の口形にもっと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。

６．２．口形推定評価

1995 年 8 月にロスア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[8]。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。

７．デモンストレーション

先に述べたプロトタイプシステムを用いて、1999 年 8 月 SIGGRAPH'99 のエマージング・テクノロジにて実際に 1 つのオリジナルストーリを作成し、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[9]。システム構成は処理用のワークステーションとして、SGI 社製 Indigo2(MIPS 10000, 123MB, IRIX6.5)、を使用し、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。投影した顔表情合成フレームレートは毎秒 8 フレーム前後で、一回のパフォーマンスで約 10 人から 30 人の観客が参加した。図 7-1 にデモンストレーションの様子を示す。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持ち、斬新かつ応用性を持つシステムであると好評を得た。また仮面に投影された顔モデルの口形もまた、同期や表情表出に対して自然であるとの意見を頂いた。

８．まとめ

本稿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について述べた。投影される顔画像は演技者の音声と同期しリアルタイムに口形状が変化する。顔表情はストーリーによって演技者が任意で変更可能である。また本システムのプロトタイプシステムとして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システムを構築し、観客とのインタラクションやオリジナルストーリのデモンストレーションを行った。

本システムはカメラトラッキング・投影・顔合成・音声分析技術を融合している。これらの技術を用いて今後、さまざまな分野への応用化を計画している[10]。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・音声分析システムを用いてフェイス・トゥ・フェイスでの多人数コミュニケーションシステム、臨場感のある電子会議システム等への応用化も検討中である。

図 7-1. デモンストレーション風景

参考文献

[1] Raskar, R. Welch, G. Cutts, M. Lake, A. Stesin, L. and Fuchs, H.: The Office of the Future : A Unified Approach to Image-Based Modeling and Spatially Immersive Displays, ACM SIGGRAPH 1998, pp179-188

[2] Cruz-Neira, Carolina, Daniel J. Sandin, and Thomas A. DeFanti.: Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE, Computer Graphics, SIGGRAPH Annual Conference Proceedings, 1993.

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

[4] A. Gelb.: Applied Optimal Estimation, The MIT Press, Cambridge, Massachusetts, 1974. Trans. on PAMI, Vol.13, No.5, pp.441-450, 1991.

[5] 森島,八木,金子,原島,谷内田,原:顔の認識・合成のための標準ソフトウェアの開発, 電子情報通信学会技術研究報告, PRMU97-282, Vol.97, No.596, pp129-136, 1998

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

[7] Morishima, S.: Modeling of Facial Expression and Emotion for Human Communication
System. Displays 17, pp.15-25, Elsevier, 1996.

[8] Morishima, S.: Better Face Communication, Visual Proc. SIGGRAPH'95, p.117, 1995

[9] Binsted, K. Nielsen, F. Pinhanez, C. Morishima, S and Yotsukura,T.: Hyper Mask: Virtual Reactive Face for Storytelling, Emerging Technologies. SIGGRAPH'99, p.186, 1999

[10]Binsted, K. Misawa, T. Morishima, S. and
Nielsen, F.: Denger Hamster 2000, Emerging Technologies. SIGGRAPH 2000, p.81, 2000

・四倉 達夫 (Tatsuo YOTSUKURA)
所属：ATR 知能映像通信研究所／成蹊大学工学部
〒619-0288 京都府相楽郡精華長光台 2-2-2
TEL:0774-95-1435 FAX: 0774-95-1408
Email: yotsu@mic.atr.co.jp
研究活動暦：平成 10 年成蹊大学工学部電気電子工学科卒。平成 12 年同大学大学院修士課程了。現在同大学大学院博士課程在学中。平成 12 年より ATR 知能映像通信研究所研修研究員。顔画像認識・合成システム、多人数コミュニケーションシステムの研究に従事。

・Kim BINSTED
所属：i-chara 株式会社
〒151-0064 東京都渋谷区上原 2-34-1
TEL:03-5465-1902 FAX:03-5465-1904
Email: kimb@i-chara.com
研究活動暦：Kim Binsted received the B.S. degree from Physics Department, McGill University of Montreal in 1991 and the M.S. degree from AI Department, University of Edinburgh in 1992.

She defended her Ph.D. thesis on "Machine humour: An implemented model of puns" prepared at AI Department, University of Edinburgh in 1996. In 1998, She joined Sony Computer Science Laboratories, Tokyo (Japan) as an associate researcher. In 2000, She joined I-Chara Inc, Tokyo as a CEO founder. Her current research interests the developing character-based social networking application for internet-capable mobile phones.

・Frank NIELSEN
所属：ソニーコンピュータサイエンス研究所
東京都品川区東五反田 3-14-13 高輪ミューズビル 3F
TEL:03-5448-4380 FAX:03-5448-4273
Email: frank@csl.sony.co.jp
研究活動暦：Frank Nielsen received the B.S. and M.S. degrees from Ecole Normale Superieure (ENS) of Lyon in 1992 and 1994, respectively. He defended his Ph. D. thesis on “Adaptive Computational Geometry” prepared at INRIA Sophia-Antipolis in 1996. As a civil servant of the University of Nice (France), he gave lectures at the engineering schools ESSI and ISIA (Ecole des Mines). In 1997, he served army as a scientific member in the computer science laboratory of Ecole 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
所属：IBM T.J. Watson Research
30 Saw Mill River Rd. (Route 9A) - Hawthorne, NY 10532
TEL: +1-(914) 945-3000 FAX:+1-(914) 784-7455
Email: pinhanez@us.ibm.com
研究活動暦：Claudio Pinhanez received the B.S. degree from Institute of Mathematics and Statistics, University of Sao Paulo in 1985 and the M.S. degree from Dept. of Computer Science, University of Sao Paulo in 1989. He defended his Ph. D. thesis on “Representation and Recognition of Action in Interactive Spaces” prepared at Media Arts & Sciences Program, MIT Media Laboratory in 1999. Currently, he joins IBM TJ Watson Research Center, NY as a research scientist.

・鉄谷 信二 (Sinji TETSUTANI)
所属：ATR 知能映像通信研究所
TEL:0774-95-1420 FAX: 0774-95-1408
Email: tetutani@mic.atr.co.jp
研究活動暦：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年、ATR 通信システム研究所に出向、臨場感表示技術に従事。平成６年、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。平成 12 年、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。

・森島 繁生 (Shigeo MORISHIMA)
所属：成蹊大学工学部／ATR 知能映像通信研究所
〒180-8633 東京都武蔵野市吉祥寺北町 3-3-1
TEL/FAX:0422-37-3726
Email: shigeo@ee.seikei.ac.jp
研究活動暦：昭和 57 年東京大学・工・電子卒。昭和 59 年同大・院・修士課程了。昭和 62 年同博士課程了。工学博士。同年成蹊大学工学部電気工学科専任講師。昭和 63 年同大学助教授、平成 9 年より TAO 共有プロジェクトサブリーダー。平成 11 年より ATR 知能映像通信研究所客員研究員、明治大学理工学部非常勤講師。平成 6 年より 1 年間トロント大学客員教授。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。平成 4 年電子情報通信学会業績賞受賞。