Institute of Electronics, Information, and Communication Engineers

総合報告

新映像技術「ダイブインントゥザムービー」

A Proposal of Innovative Entertainment System "Dive Into the Movie"

森島繁生　八木康史　中村　哲　伊勢史郎　向川康博
槇原　靖　間下以大　近藤一晃　榎本成悟　川本真一
四倉達夫　池田雄介　前島謙宣　久保尋之

abstract

映像コンテンツの全く新しい実現形態として，観客自身が映画等の登場人物となり，時には友人や家族と一緒にこの作品を鑑賞することによって，自身がストーリーへ深く没入し，かつてない感動を覚えたり，時にはヒロイズムに浸ることを実現可能とする技術「ダイブインントゥザムービー」について本稿で解説する．この実現には，観客に全く負担をかけることなく本人そっくりの個性を有する登場人物を自動生成する技術と，自ら映像中のストーリーに参加しているという感覚を満足するためのキャラクタ合成のクオリティ，映像シーンの環境に没入していると錯覚させる高品質な映像・音響再現技術及びその収録技術が，観客の感動の強さを決定する重要な要素となる．

2005年の愛・地球博にて実証実験を行った「フューチャーキャスト」に端を発するこの技術は，ハードウェアの進歩と2007年にスタートした文部科学省の支援による科学技術振興調整費プロジェクトの実施によって，格段の進歩を遂げた．その結果，様々なバリエーションの観客の個性を全自動・短時間でストレスなくモデル化することが可能となり，また作品の中でリアルタイム合成されるキャラクタの顔と全身，声に各人の個性を忠実に反映することが可能となった．また，同時に役者が感じた音場・視点で1人称的にコンテンツへの没入感を体感することを可能にするシステムを同時に実現した．

キーワード：高速頭部形状モデリング，個性ある表情合成，肌の質感計測，肌のリアルタイムレンダリング，歩容計測，歩容データベース，歩容画像合成，音声配役システム，声質類似性尺度，三次元音場収録・再現，主観視点全方位映像収録・呈示

1. はじめに

観客全員が映画の登場人物にふんし，映像コンテンツ中でお互いが活躍する姿を家族や友人とともに鑑賞できる．あるいは，好みの俳優と自分自身が共演する映像作品を簡単に制作できる．更に主人公の位置から見える風景や音場を1人称的に体感することができる．このような新しいエンタテインメントの実現形態を生み出すことが，「ダイブインントゥザムービー」のコンセプトである．

このような映像作品は，生理評価実験結果から，映像を鑑賞する者の作品やストーリーへの没入感を高めることが明らかになっており，特に，自分自身や知り合いの姿をスクリーン上で発見した瞬間に急激に感情の高まりが検出され，気分が高揚する効果が確認されている[1]．

森島繁生　正員　早稲田大学理工学術院先進理工学部
E-mail　shigeo@waseda.jp
八木康史　向川康博　正員　大阪大学産業科学研究所
中村　哲　榎本成悟　正員　独立行政法人情報通信研究機構けいはんな研究所
伊勢史郎　京都大学大学院工学研究科建築学専攻
槇原　靖　大阪大学産業科学研究所
間下以大　正員　大阪大学サイバーメディアセンター
近藤一晃　正員　京都大学学術情報メディアセンター
川本真一　正員　北陸先端科学技術大学院大学情報科学研究科
四倉達夫　正員　(株)オー・エル・エムデジタル研究開発部門
池田雄介　独立行政法人情報通信研究機構けいはんな研究所
前島謙宣　正員　早稲田大学IT研究機構
久保尋之　学生員　早稲田大学大学院先進理工学研究科
Shigeo MORISHIMA, Member (School of Science and Engineering, Waseda University, Tokyo, 169-8555 Japan), Yasushi YAGI, Yasuhiro MUKAIGAWA, Members, Yasushi MAKIHARA, Nonmember (The Institute of Scientific and Industrial Research, Osaka University, Ibaraki-shi, 567-0047 Japan), Satoshi NAKAMURA, Seigo ENOMOTO Members, Yusuke IKEDA, Nonmember (Keihanna Research Laboratories, National Institute of Information and Communications Technology, Kyoto-fu, 619-0289 Japan), Sirou ISE, Nonmember (Graduate School of Engineering, Kyoto University, Kyoto-shi, 615-8540 Japan), Tomohiro MASHITA, Member (Cybermedia Center, Osaka University, Toyonaka-shi, 560-0043 Japan), Kazuaki KONDO, Member (Academic Center for Computing and Media Studies, Kyoto University, Kyoto-shi, 606-8501 Japan), Shinichi KAWAMOTO, Member (School of Information Science, Japan Advanced Institute of Science and Technology, Ishikawa-ken, 923-1292 Japan), Tatsuo YOTSUKURA, Member (Research & Development Division, OLM Digital, Inc., Tokyo, 154-0023 Japan), Akinobu MAEJIMA, Member (Information Technology Research Organization, Waseda University, Tokyo, 169-8555 Japan), and Hiroyuki KUBO, Student Member (Graduate School of Advanced Science and Engineering, Waseda University, Tokyo, 169-8555 Japan).
電子情報通信学会誌　Vol.94 No.3 pp.250-268 2011年3月
©電子情報通信学会 2011

NII-Electronic Library Service

Institute of Electronics, Information, and Communication Engineers

また，老人やハンディキャップを有する人物を勇気付けるという効果も実際に報告され，次世代の映像コンテンツの表現形態の一つとして，方向性を示唆するものである．

2005年の愛・地球博で実現され，その後，長崎のハウステンボスで常設展示館となっている「フューチャーキャスト」は，このコンセプトを実現する世界初のシステムであり[2]，万博期間6か月の体験者数163万人という記録とハウステンボスでの今も変わらぬ人気ぶりは，この革新的エンタテインメントがいかに人々の感動を高めることに成功しているかを実証するものである．

更にこの技術水準を高めるべく，2006年8月～2009年3月までの約2年半の期間をかけて，文部科学省の科学技術振興調整費の支援のもと，筆者らは研究開発プロジェクトを実施した．

その技術達成課題は，まず第1にいかに短時間で人物の個人性特徴を忠実に再現したCGキャラクタのリアルタイム合成を実現するかであった．顔形状，顔表情，肌の質感，体格，歩き方（歩容），声質各々の個人性表現と自動特徴抽出に焦点を当て，自分自身の姿を，いかに容易に違和感なく映像作品中に見いだすことができるかという評価基準のもとプロトタイプシステム実装を行い評価実験を行った．その結果，通常手作業で数か月を要する人物の忠実な個人性を有するモデリング手法を，簡単で被験者に負担のない計測方法と高速な処理アルゴリズムにより，撮影から上映可能なモデル完成まで，わずか10分程度の時間で達成し，また映像中に自己を発見する自己認識率の点では，87%を記録することができた．これは，2005年の「フューチャーキャスト」が自己認識率57%であったのと比較して格段にシステムの性能が向上したことを意味する．

プロジェクト第2の目標は，主人公視点の映像と三次元の音場を忠実に収録し，空間にありのままを再現するシステムの実現である．この成果として高品位な全方位映像を三次元音場と同時に記録するシステムを実現し，また同時に四方に配置した平面スクリーンとスピーカアレーによって，主人公を取り囲む環境映像と音場を，ユーザにディスプレイするシステムを実際に構築した．

本稿では，この「ダイブイントゥザムービー」プロジェクトで実現された各要素技術を中心に解説し，革新的エンタテインメント実現に不可欠な要素技術について論ずる．

2. ダイブイントゥザムービーのための顔画像処理技術

顔の特徴は個人性を反映したキャラクタを表現する上で，最重要な要素であり，特徴表現することが困難な研究ターゲットでもある．本章では，個人の頭部の立体形状，表情を反映可能な三次元頭部モデリング手法，個人の肌の質感計測とリアルタイムレンダリング手法について述べる．

2.1 三次元頭部形状モデリング

ダイブイントゥザムービーシステムにおいて，低コストで体験者に忠実なキャラクタの頭部モデルを短時間自動生成する技術は，極めて重要な要素技術の一つとして位置付けられている．

一般的に人物の頭部CGモデルの作成の際は，360度全周囲の立体形状が計測可能なレンジスキャナ等を用い，三次元の立体形状を取得して，それに汎用の三次元メッシュモデルを整合することによって，表情制御可能なキャラクタ頭部モデルを生成している．しかし，この手法では一般に①計測時間が長い，②計測中に体験者が一定時間同じ姿勢を保つ必要がある，③原理的に頭髪に覆われた後頭部は正確な形状は測れない，④レンジスキャナ装置が極めて高価である，という問題点がある．

よって比較的短時間でかつ体験者の負担を軽減するため，確実に個人性を反映せねばならない顔前面は専用のレンジスキャナにより計測した三次元形状を用い，比較的精度が要求されない後頭部はテンプレートメッシュにより形状を補完することで，個人の頭部モデルを生成する．

2.1.1 モデリング手法

三次元頭部モデリング手法の流れを図1に示す．システムへの入力は，(a)顔テクスチャと(b)三次元顔形状及び(c)テンプレートメッシュである．三次元顔形状はデプスマップ形式であるため，テクスチャとピクセル単位の対応付け情報を持つ．よってテクスチャ中の任意の座標を指定するとその奥行き座標を取得できる．

頭部モデル生成の流れは次のとおりである．(d)顔テクスチャから，目，まゆ，鼻，口，顔輪郭上の149点の特徴点を自動検出する．次に，(e)検出された特徴点とそれに対応するテンプレートメッシュ上の頂点の対応関係を用いて，普遍型クリギングによる回帰に基づきテンプレートメッシュをテクスチャ中の顔に対して整合する．(f)テクスチャに対して整合されたテンプレートメッシュの顔領域の頂点座標を，そのuv座標に対応する個人のレンジスキャンデータの三次元座標で置き換える．これによりテンプレートメッシュの顔領域に個人の特徴が反映される．次に，(g)結果のメッシュの顔領域のみの姿勢と大きさを，テンプレートメッシュのそれと合うように正規化する．結果として，顔領域とそれ以外の領域との間につなぎ目が見られるものの，後頭部の形状が付加された仮の頭部モデルが出来上がる．こうして得られた(i)メッシュの頂点をマーカとし，後述する最適化アフィン変換に基づくフィッティング手法により再

NII-Electronic Library Service

Institute of Electronics, Information, and Communication Engineers

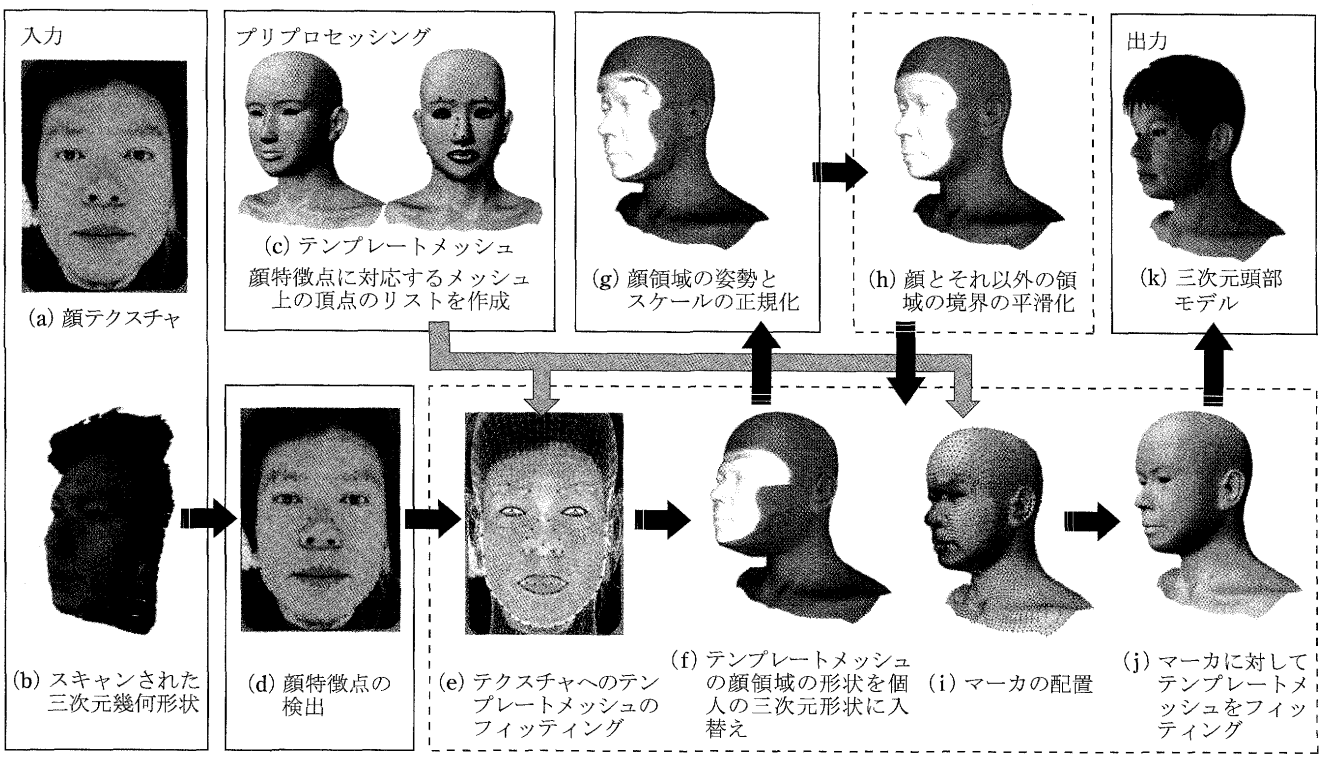


図1　三次元頭部モデリングの流れ

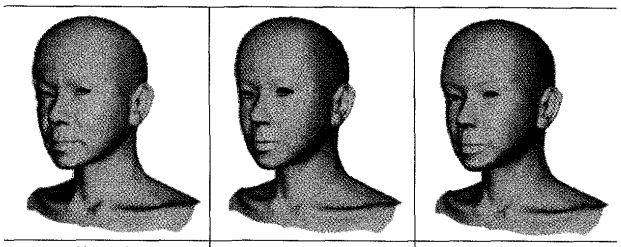
度(j)テンプレートメッシュをフィッティングする．このフィッティングにより顔領域とそれ以外の領域との間に(k)つなぎ目のない個人の三次元頭部モデルを生成することが可能となる[3]．

2.1.2　モデリング結果

モデリング手法の有効性を評価するために，この手法をC++，GNU Scientific Library 1.8及び，L-BFGS-B Library[4]を用いて実装し，従来Allenらが提案しているフィッティング手法[5]（以下，Allen'03と呼ぶ）との比較実験を行った．

具体的には，両手法により100名分のレンジスキャンデータから頭部モデルを生成し，その平均処理時間を計測するとともに，うち3例についてフィッティングに用いたマーカとの誤差を計測した．実験環境は全てIntel Core (TM) 2 Duo CPU T9600 2.8 GHz，3 GByteメモリのマシンにて実施し，テンプレートメッシュとして，頂点数6,586点，ポリゴン数13,036個の三角形ポリゴンにより構成される頭部のメッシュモデルを用いた．

図2に評価結果を示す．(a)がマーカ間をつなぎサーフェスで表現したフィッティングターゲット，(b)がAllen'03により生成された頭部モデル，(c)が提案手法により生成された頭部モデルである．フィッティングターゲットと各手法により生成されたモデルの顔領域内のマーカと頭部モデルの全頂点間の二乗平均平方根誤差（RMS Error : Root-Mean-Squared-Error）は，それぞれ0.025と0.019であり，モデルの両目間距離に対する



	(b) Allen'03	(c) 提案手法
二乗平均平方根誤差	0.025(3.2%)	0.019(2.4%)
平均処理時間(s)	377.7	10.4

(a) ターゲット形状

図2　従来手法と提案手法の性能比較

比で表した場合3.2%と2.4%であった．

また，両手法の平均処理時間はそれぞれ，377.7秒と10.4秒であった．この結果から，提案手法はAllen'03よりも高いフィッティング精度を実現しながら，同時に36.3倍高速であることが分かった．

図3は最上行がテンプレートメッシュを表し，最左列が個人のレンジスキャンデータ，2〜4列目が各テンプレートメッシュから生成された顔モデル，5列〜8列目が頭髪モデルの変更による印象の違いを表している．

2.1.3　顔写真1枚からの顔三次元形状モデリング

最新の技術として，レンジスキャナを用いることなく正面顔写真1枚から三次元顔モデルを高速自動生成する手法の研究[6]を進めている．提案手法は，1〜2秒で本人の三次元形状と比較して頂点レベルで約2 mmの平均

NII-Electronic Library Service

Institute of Electronics, Information, and Communication Engineers

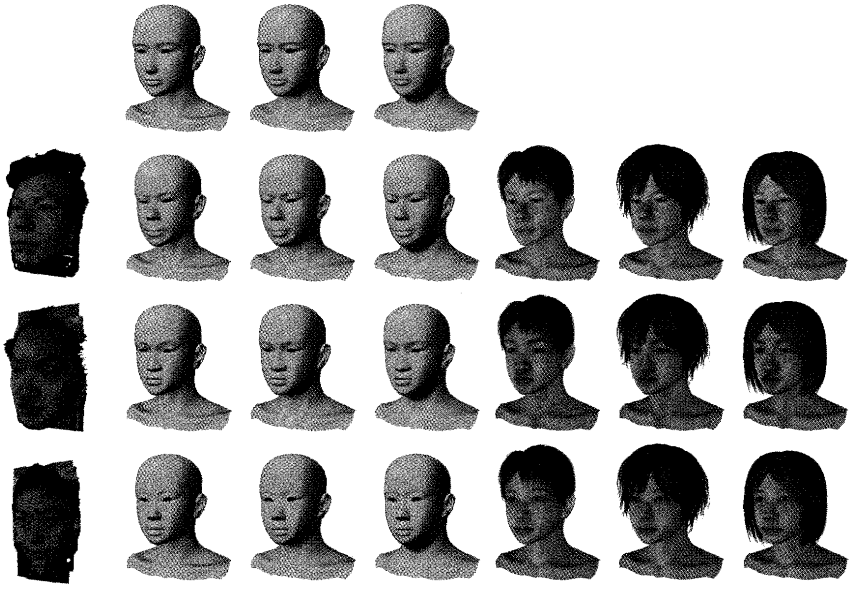


図3　異なるテンプレートメッシュを用いて生成された頭部モデルと最終的なモデリング結果

図4　顔写真1枚から生成された三次元顔モデル

誤差精度を持つ三次元顔モデルの自動生成を実現している（図4）．具体的には，約1,200人分の顔の立体形状のデータベースから学習される統計的な顔の変形モデル（以後，顔変形モデルと呼ぶ）を用い，顔画像から検出される特徴点と顔変形モデルを二次元平面上に投影したときの頂点との間の誤差を最小化し，顔変形モデルを個人の顔に合うように変形することで顔の三次元形状が推定されるという仕組みである．この最小化の際に，頭部の形状分布を表す混合ガウス分布に対する顔らしさのゆう度を最大化するように問題を設定することで，推定される顔形状の自然さを保持するような工夫をしている．これにより，提案手法を用いれば，極めて低コストで実用に耐え得る精度の三次元顔モデルを自動生成することが可能である．

この技術によって，顔写真1枚さえあれば，表情制御可能な顔メッシュモデルを自動生成可能であり，ホームユースとして，Web上の写真のみから，登場人物をカスタマイズすることができる．またプロフェッショナルユースとしても，従来手作業で行われていた特定人物のキャラクタ生成作業の大幅な効率化が実現できる．

2.2　個人特徴を反映する表情自動合成

三次元CGで作成されたキャラクタに対して物理的な破綻がなく，リアルで表情豊かなアニメーションを付加する作業は，熟練のクリエイタでさえ複雑で手間のかかる作業である．更に個人性特徴の実現となれば，試行錯誤によるパラメータチューニングが必要不可欠となり，作業は更に困難を伴う．三次元CGキャラクタのリアルな表情自動合成は，ニーズの高い技術である．

2.2.1　表情筋パラメータを用いた表情合成

本提案手法は基本的に，ある人物が指定された表情を表出している動画像と本人の無表情のレンジデータを入力とし，表情筋収縮パラメータのフィッティングと個人

NII-Electronic Library Service

Institute of Electronics, Information, and Communication Engineers

性適応により任意の表情表出に対して個人性を反映しリアルに表情アニメーションを生成することができる．これにより，表情アニメーションのベースとなる，表情筋収縮パラメータの時間遷移データ（ベースアニメーション）が取得される[7],[8]．このときマーカの貼付など不要で，指定された表情表出を1回だけ行えばよく，体験者への負担は極めて少ない．

取得されたベースアニメーションに対して，体験者の任意な特徴的要素を付加することで，より“その人らしい”表情アニメーション作成する．また，表情筋を介することで，表情表現のパラメータを統一的に扱い，ストーリーのシナリオとして位置付けできる点が特徴である．

まず最初に別の人物のフェースキャプチャ結果に基づき，表情筋パラメータの推定を行い，クリエイタの編集・手修正を経て決定された各表情筋の強度をあらかじめシナリオと定義する．

体験者の無表情から笑顔への表情変化の動画像を取得する．この動画像中で，初期無表情フレームにおける89点の静的顔特徴点自動抽出結果に基づき，標準ワイヤフレームの自動フィッティングを動画の全フレームに対して行い，このワイヤフレームの各頂点を新たな追跡ターゲット（特徴点）と定める．表情変化に伴う特徴点追跡をLucas/Kanadeオプティカルフローアルゴリズム[9]に基づいて行い，測定された頂点の移動量に対して主成分分析を適用することで最適な表情筋配置を推定し，個人顔モデルにマッチした表情筋配置を決定する[10]．

2.2.2 脂肪層を考慮した表情合成

表情は表情筋そのものの性質だけでなく，顔面の脂肪層の厚みによっても大きく左右される．そこで，MRIより取得した実際の構造を筋肉モデルに反映させ，表情の個性を反映可能な表情合成手法を提案する．

脂肪層の体積は，MRIの各スライス画像の輝度情報からしきい値処理により切り出された脂肪層の面積の重ね合わせで計算する．同時に頭部の体積を計算し，頭部体積と脂肪層体積との比を分割領域ごとに求め，その値を個人のパラメータとして定義する．このパラメータを基に脂肪層の厚みを増減させて表情筋シミュレーションを実行することで表情合成を行う．

脂肪層の厚みを考慮した表情合成結果と本人の表情との比較を図5に示す．シミュレータによって各筋肉に付加された力は同一であるが，脂肪層の厚みの違いによって，(a)と(b)との表情合成結果の間で明らかに頬顎部位において印象が異なることが確認できる[11]．

2.3 肌の質感モデリング

肌の質感は，個人性表現要素として重要であり，年齢，性別，環境，疲労，コンディションなど，様々な要因で変化する特徴である．また肌の透明感は，キャラクタのリアリティを左右する重要な要素でもあり，その特徴を捉えることは困難で，かつレンダリングに要する計算量は一般的に極めて膨大である．

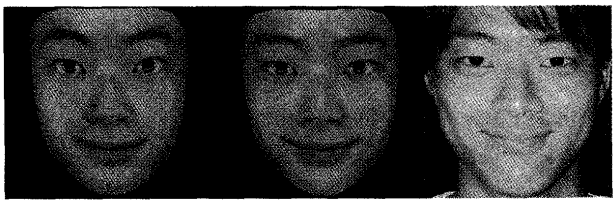
2.3.1 肌で生じる反射光と散乱光

人間の肌の個人性は単なる色やてかり具合の違いだけではなく，透明感の違いによっても生じる．例えば，子供の肌は透き通るように見え，化粧している肌は滑らかに見える．これは，肌が光を内部に通す半透明な性質を持ち，表面での光の反射や内部での散乱の様子が個人によって異なるからである．

肌は図6に示すような多層構造をしている．各層は半透明であり，肌表面への入射光の一部は下位の層にまでたどり着く．肌の内部に入り込まない光は，極めて表面に近い層で反射されるので反射光と呼ぶ．この反射光は，更に拡散反射と鏡面反射に分けることができる．一方，肌の内部に入り込んだ光は，複数の層で反射を繰り返した後，入射点とは異なる位置から出射するため散乱光（表面下散乱）と呼ぶ．

2.3.2 質感の計測と解析

人間の肌の反射・散乱特性を計測するために，図7に示す計測システムを構築した[12]．この計測システムは箱の1面に顔を入れる穴が開いており，その正面と右側に2台の高感度カメラを，正面と左右に3台のプロジェ



(a) 他人の脂肪層　(b) 本人の脂肪層　(c) 本人の画像

図5 脂肪層の厚みを考慮した表情合成結果

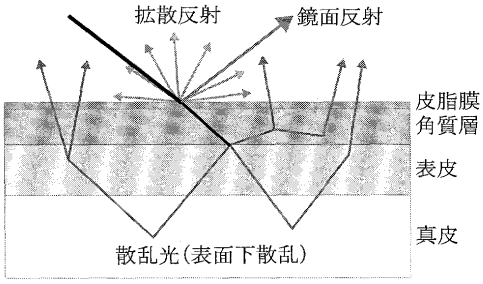


図6 人間の肌の多層構造

NII-Electronic Library Service

Institute of Electronics, Information, and Communication Engineers

クタを設置している．プロジェクタから様々なパターン光を投影してカメラで撮影することで，人間の顔の幾何情報と反射特性を計測できる．本システムによる計測時間は約30秒である．

肌の質感を定量的に評価するために，肌の見え方を成分ごとに分離し，それぞれの成分の強度を調べる．反射光と散乱光を分離するために，本研究では高周波パターンを投影する[13]．顔の正面に配置したプロジェクタから，細かいしま状の繰返しパターン（高周波パターン）を投影し，正面のカメラで撮影する．高周波パターンを空間的にずらしながら投影することで物体上の各点は照明・非照明の状態に変化する．照明時の各点への入射光は反射光成分と散乱光成分を併せ持ち，非照明時は散乱光成分のみであるため，この差分に基づいて反射光成分と散乱光成分を分離する．

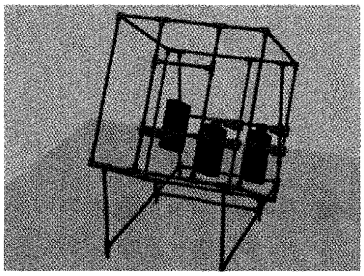
図8に，反射光と散乱光を分離した結果を示す．なお，被験者には円形マーカを貼付したサングラスを装着してもらい，プロジェクタ光から目を守ると同時に，両目との相対的な位置関係から額と頬の位置を特定している．反射光成分は全体的に青色であり，散乱光成分は全体的に赤色である．これは，波長の短い青い光が肌の内部に入りにくく，肌の内部に入り込んだ散乱光が血液中のヘモグロビン色素の影響を受けるからと考えられる．

次に，反射光を拡散反射と鏡面反射に分離する．本研究では，観測強度の視線依存性の違いを利用して両成分を分離する．鏡面反射は視線方向に依存するが，拡散反射は視線方向に依存しない．そこで，顔を正面方向と斜め方向から2台のカメラで撮影する．プロジェクタからグレイコードパターンを投影することで，2台のカメラ間で画素を対応付ける．これによって斜め方向から見た顔画像を正面方向に変換できる．鏡面反射は視線方向に依存するため，実際の正面画像と正面に変換した画像では鏡面反射の位置が異なる．この差分に基づいて鏡面反射成分を分離する．

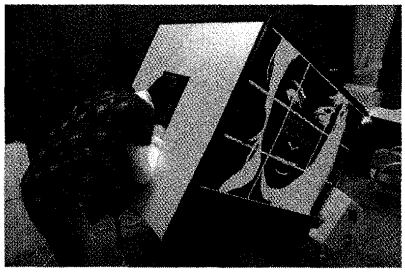
図9に反射光を拡散反射成分と鏡面反射成分に分離した結果を示す．肌のてかりが，鏡面反射成分として鼻や唇の周辺で強く観測されていることが分かる．

このように，拡散反射・鏡面反射・散乱光の各成分に分離することで，額や頬での各成分の強度比を定量化することができ，CGにより質感を再現することができる．更に，図10のように，各成分の重みを変化させて再合成することにより，てかりや透明感を誇張して見せるなどの応用も可能である．

今後の課題として，個人に忠実な質感の再現だけでなく，年齢や性別に応じた代表的な質感モデルの構築があげられる．筆者らは，1,000名を超える大規模データベースを用いて肌の質感を統計的に解析している[12]．

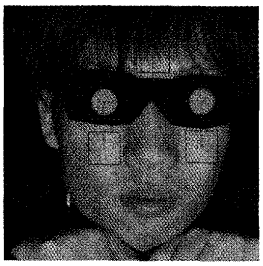


(a) システムの構成

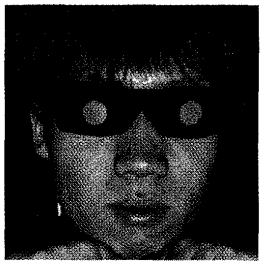


(b) 計測の様子

図7　肌の質感計測システム



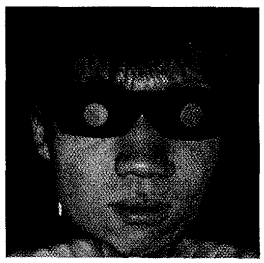
(a) 通常照明



(b) 反射光成分

(c) 散乱光成分

図8　表面下散乱の強度解析



(a) 拡散反射

(b) 鏡面反射

図9　拡散反射と鏡面反射の分離

NII-Electronic Library Service

Institute of Electronics, Information, and Communication Engineers

図10　肌の質感の変化　左から順に，入力画像と，それを乾燥肌，脂っぽい肌，透明感のある肌，不透明感な肌に変化させた見え方．

また，肌の異方性散乱と不均一性の表現[14]や，化粧の効果の再現[15]にも取り組んでいる．

2.4　肌の質感リアルタイムレンダリング

三次元CGにおいて，様々な物質の反射特性をモデル化することはフォトリアリスティックなレンダリングを実現する上で極めて重要といえる．

しかし，人間の肌や大理石に代表される半透明な物体は，入射光が物体内部に進入し散乱を繰り返す．そのため，物理的に正確なモデル化には，大局照明モデルを導入せざるを得ず，結果的にリアルタイムなレンダリングには，計算量的にほど遠いのが現状である．

表面下散乱（SSS：Sub Surface Scattering）を考慮しない古典的なランベルト則と，SSSを考慮したレンダリング結果とを比較したところ，物体表面の起伏に富んだ点においてSSSによる影響は特に顕著であることが分かった．そこで，物体表面の起伏の強さを表す指標として，曲率を導入する．曲率は物体表面上で局所的に決定されるパラメータであるため，局所照明モデルが適用可能であり，高速なレンダリングが実現される．

まず与えられた物体の物質パラメータを半径rの球に適用し，単一の白色な平行光源下で球表面の放射輝度分布を推定する．半径rの球面の曲率は$1/r$で表されるため，様々な半径の球面に対する放射輝度分布を調査することにより，曲率に依存する反射関数（CDRF）を得る．

次に，あらかじめ曲率が計算されたポリゴンモデルに対しCDRFを適用し，レンダリングする．ただし，CDRFはCPUによって2Dのルックアップテーブルとして作成し，テクスチャとしてあらかじめグラフィックスメモリに転送しておく．描画時にはピクセルシェーダによるテーブル参照のみで実行可能となり，リアルタイムに表面下散乱の影響を考慮したレンダリング（図11）が可能となった[16]．

3.　ダイブインントゥザムービーのための歩容画像処理技術

人間の体形や歩容動作には個人性があることが知られ

(a) 直接反射のみ　　(b) 提案手法

図11　曲率を考慮した半透明物体の描画結果（© コーエーテクモゲームス）

ており，この個人性により遠方からでも家族や友人を区別できた経験を多くの人が持っている．本研究ではこれらの体形・歩容特徴をオンラインで計測して，CGキャラクタへ反映させることを目的とする．

歩容個性計測の方法としては，ウェアラブルセンサによる手法[17]や人体モデルのフィッティングによる方法[18]も考えられるが，来場者によるセンサの装着時間がかかること，モデルフィッティングの計算コストが高いことから，オンライン計測には不向きである．そこで本研究では，来場者が自然に歩行する様子をカメラで撮影し，計算コストの低いシルエットベースの解析によって体形や歩容特徴をオンライン計測するものとする．

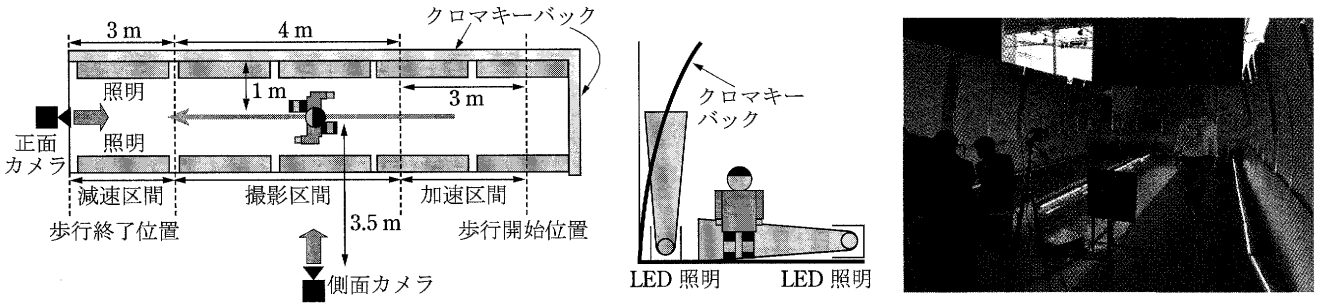
3.1　歩容画像データベースの構築

3.1.1　オンライン歩容個性計測システムの構築

来場者の歩容個性を30秒程度で計測することを目的として，側面と正面の同期カメラ2台，撮影用の計算機1台，歩行者のシルエット抽出を容易にする緑色の壁面とカーペット，及びそれらに対するLED照明からなるオンライン歩容個性計測システムを構築した．歩行路の長さは10mに設定し，そのうち，歩き始めからの3mを安定歩行に至るまでの加速区間，歩き終わりの3mを減速区間として，実質4mの歩行を撮影するものとした（図12）．

NII-Electronic Library Service

Institute of Electronics, Information, and Communication Engineers



(a) 真上図　(b) 真横図　(c) 撮影の様子

図 12　オンライン歩容個性計測システムの概要と撮影の様子

3.1.2　オンライン歩容個性計測手法の開発

本項では歩容個性の計測手法の概要について述べる[19]．対象とする歩容個性は，オンライン計測装置によって得られた歩容映像から計測可能で，かつ，個人間での分散が大きく，映像中で人が直観的に認識しやすいことが必要である．そこで，体形に関する静的特徴量として，身長・体の幅・厚みを，歩容動作中に見られる動的特徴量として，歩行速度・歩行周期・歩幅・左右の腕の前後の振り・足の運びにかかる時間の左右非対称性・猫背具合を計測する．

まず，前処理として，映像から背景差分によって歩容シルエットを抽出する．次に，歩容の周期性に着目して，足が開いた両足接地状態（DSP）や足が閉じた片足接地状態（SSP）に対応する基準フレームを，垂直中心軸周りの二次モーメントの最大・最小フレームとして各々検出する．

（1）　静的特徴量の計測

事前にカメラと歩行者の位置関係を校正しておくことで，SSP のシルエット画像上の画素単位の身長，体の幅，体の厚みから実際の長さに変換することで計測する．

（2）　動的特徴量の計測

（a）　歩行速度・周期・歩幅

計測開始・終了時点での人物中心位置の差と経過時間により歩行速度を計測する．また，左右 1 歩ずつの合計 2 歩を一周期として，シルエット画像列の正規化自己相関により歩行周期を検出し[20]，歩行速度・周期から歩幅を推定する．

（b）　腕の振り

SSP における腹・背中側の体のラインを各々抽出しておき，SSP から DSP までに各ラインの前後の上半身領域においてシルエットが掃引する面積を，左右の腕の前方・後方への振りとして個別に計測する．また，透視投影の原理に基づいて，足下点がより下側に映る脚を手前，もう一方を奥側の脚と判定し，進行方向や一般的に手足が逆位相で動くという事実に基づいて，腕の左右を判断する．

（c）　足の運びの左右非対称性

隣接 DSP 間のフレーム数が各脚の運びにかかる時間を表しており，その差を脚の運びの左右非対称性の尺度とする．

（d）　猫背

SSP のシルエット画像から，背中側の体の曲線を抽出し，その曲線に対する直線当てはめによって傾きを抽出し，更に曲線から k 曲率を計測することで，それぞれ猫背度合いの尺度とする．

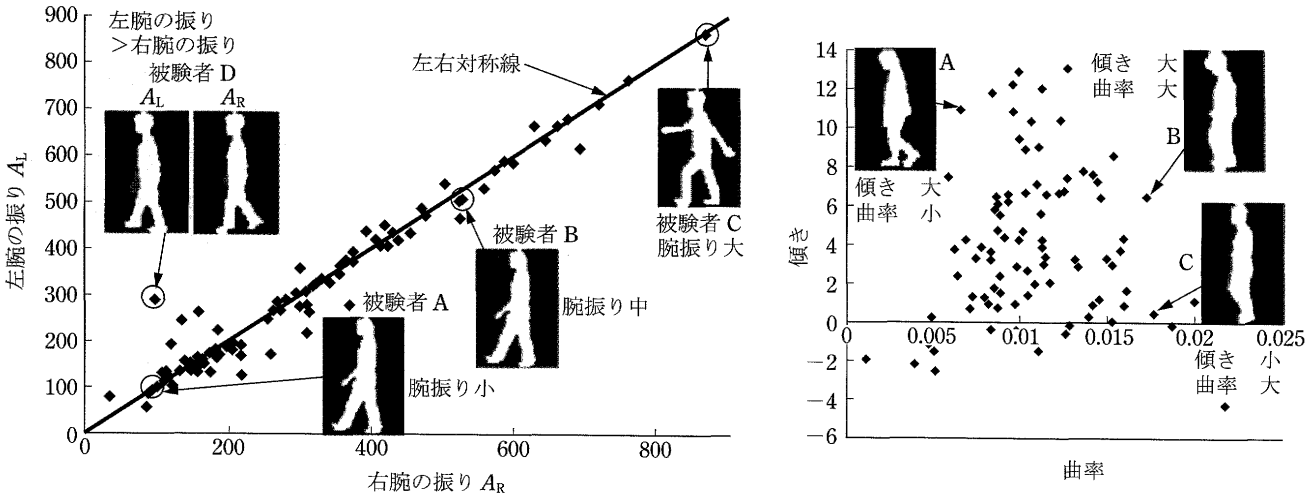
これらの特徴量の有用性を確認するために，多視点同期歩容計測システム[21]により収集した歩容データベース[22]からサンプリングした 100 人の被験者から歩容個性を抽出した．動的特徴量の一部の歩容特徴の分布を図 13 に示す．これより個人間の分散が十分に存在することを確認した．

3.2　個性ある歩容画像合成

3.2.1　個人性を有する全身モデルの自動生成

オンライン計測によって抽出される静的特徴量に基づいて，キャラクタモデルのサイズ（身長，体の太さ，首の太さ，あごのサイズ，皮下脂肪層の厚み）を決定する．図 14 に，既に述べた頭部モデリング技術を応用して，異なる首の太さの後頭部モデルに対して，同一の個人の計測された顔のレンジデータに基づいて生成された頭部モデルを示す．また，全身のキャラクタモデルは，典型的な九つの体形をクリエイタにデザインしてもらい，それらの補間によって中間体形モデルを生成する手法を提案した．ただし，すべてのモデルは同一のポリゴン構造と頂点構成を持ち，ボーンの配置とスキニングウェイトは，個別のサイズのキャラクタに最適化されている場合，それらの補間によって求めた任意の体形のキャラクタモデルは，動作パラメータとして骨格の角度データを与えた場合，破綻せずに動作することが確認できた．

NII-Electronic Library Service



(a) 腕の前振り　　(b) 猫背度合い

図 13　動的特徴量の分布

図 14　異なる体形に対応した頭部モデル　　左ほどスリムな体形，右ほどファットな体形に対応する頭部モデルの生成結果．体形は静的特徴量に基づいて決定される．

(a) 被験者 A

(b) 被験者 A の合成結果

(c) 被験者 B

(d) 被験者 B の合成結果

(e) 被験者 C

(f) 被験者 C の合成結果

図 15　シルエット画像と CG 合成結果

NII-Electronic Library Service

Institute of Electronics, Information, and Communication Engineers

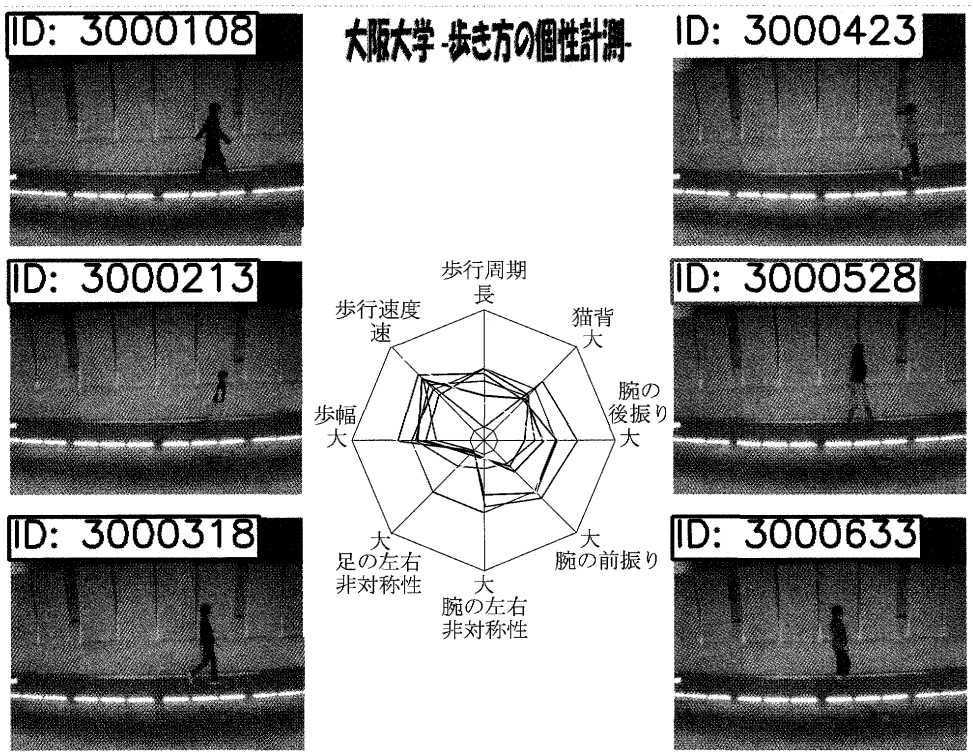


図 16　複数被験者間の歩容個性計測の比較画面　各被験者 ID の枠線の色と中央のレーダグラフの色とが対応している．中央のレーダグラフの各項目は，上から時計回りに，歩行周期，猫背，腕の後振り，腕の前振り，腕振りの左右非対称性，足の運びの左右非対称性，歩幅，歩行速度の大小を表す．

3.2.2　歩容動作モデルの自動カスタマイズ

ここでは，計測された動的特徴量に基づいて CG キャラクタの歩容動作を再合成することを目的とする[23]．

まず，歩行機上での様々な歩き方（歩行速度は 4 km/h で固定）のモーションキャプチャデータを収録し，2 歩行周期を切り出して標準歩容動作モデルとする．次に標準歩容動作モデルに対応する CG モデルを作成してシルエット映像を合成することで，オンライン歩容個性計測と同様の手法によって，静的・動的特徴量を抽出しておく．

これにより，オンライン計測した特徴量に対して，最も特徴が近くなるように CG モデルをブレンディングすることで，被験者の身体・歩容動作を CG キャラクタ上で再現する．基本動作のフレーム長は，その歩容特徴によって異なるため，これらのモーションのブレンディングには，時間軸正規化法を用いる[24]．異なる 3 人の歩容シルエット映像とそこから抽出された歩容個性に基づいて再現された CG キャラクタの合成映像を図 15 に示す．結果として，各被験者の腕の振りや猫背等の特徴をよく捕らえていることが分かる．

3.3　個人性検証実験

日本科学未来館において実施したアウトリーチ活動において，オンライン歩容個性計測システムを用いて検証実験を行った．三日間の合計で約 1,600 人の来場者データを収集し，オンライン体形・歩容計測及びブレンドモデルの合成が行えることを確認した．歩容個性計測については，撮影を含めて一人当り 30～40 秒程度で行うことができ，家族や友人間での歩容個性の比較（図 16）をオンラインで提示することができた．

4.　ダイブイントゥザムービーのための音声処理技術

映像コンテンツに登場する人物のキャラクタを演出する上で登場人物が話す音声（シナリオ音声）は重要な役割を担っている．本章では，利用者の音質を映像コンテンツの登場人物に反映させる技術の実現のため，多量の話者のシナリオ音声をあらかじめ蓄積しておき，利用者の音声に近い音色の音声を生成する方法と，生成した音声を映像と同期して出力するシステムについて紹介する．

4.1　声質の類似性評価

あらかじめ用意された多量の話者のシナリオ音声から，利用者の音声に近い音色の音声を選択するには，音声の個人性の類似性評価技術が必要となる．この課題を解決するため，個人性の知覚に関係があるとされる複数の特徴量の重み付き線形結合により類似性を評価することで，利用者の音声に対して知覚的に類似するシナリオ

NII-Electronic Library Service

Institute of Electronics, Information, and Communication Engineers

音声をデータベースから選択する技術を開発した．各音響特徴量の結合重みは，人により付与された知覚的類似度との順位相関係数を用いて最適化した．

音響特徴量としては，個人性の知覚に関係があると報告されている下記の8種類を採用した．

MFCC：Mel Frequency Cepstral Coefficient
（MFCC 12次＋ΔMFCC 12次＋Δパワー1次の計25次元）
CepH：35次以上のSTRAIGHTケプストラム[25],[26]
Cep1：第1次のSTRAIGHTケプストラム
Spectrum：2.6 kHz以上のスペクトル
Ap：2 kHz以下のSTRAIGHTの非周期性指標 STRAIGHT-Ap
F0：基本周波数
Formant：1次から4次までのホルマント
SpecSlope：3 kHz以下の対数スペクトル傾斜

各特徴量間の距離尺度はDynamic Time Warping（DTW）距離とGaussian Mixture Model（GMM）ゆう度を用いた．各特徴量の結合重みは，本研究において評価基準としたSpearmanの順位相関係数を増大させるように最急降下法によって最適化した．なお，本実験では事前に知覚的類似度を付与した36名分の音声を用いた．

知覚的類似度推定精度の評価実験結果を図17に示す．横軸に音響特徴量を示し，提案法は「Combi」としている．縦軸は順位相関係数であり，提案法は単独の音響特徴量を用いて推定した場合に比べ，高い相関係数を得ていることが分かる．またDTWはGMMより高い推定精度であることが分かる．DTWはイントネーションのような大局的な音響特徴量の時間構造を考慮しているため，より人の類似度判断に近い結果が得られたと考えられる．単独の音響特徴量を用いた結果の中で，最も推定精度が高かったのはDTW距離を用いたF0であるが，提案法は平均値で0.11上回っており，推定精度の向上が確認できた[27]．

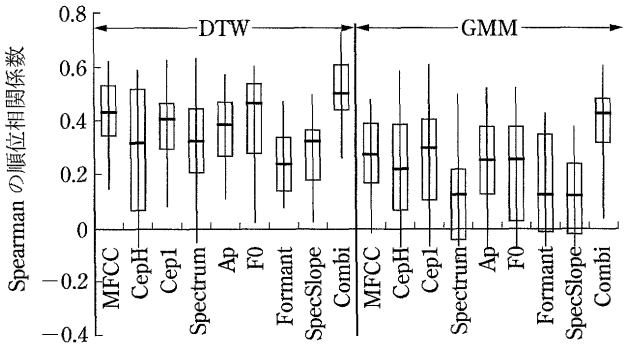


図17　知覚的類似度の推定率の比較

4.2　音声配役システムの構築

4.1で述べた類似性評価技術に基づいて，多量の話者のシナリオ音声から選択した音声を映像と同期して出力するシステムを構築する．映像との同期が確保された高品質なシナリオ音声データベースから利用者の声に類似する話者を選択し，同期再生する方式を採用することで，劣化のない高品質なシナリオ音声と高い映像同期精度を安定して確保できるシステムが実現可能になる．

4.2.1　シナリオ音声データベース

既存の映像コンテンツに合わせてシナリオ音声を出力する応用を想定する．同じ声質を維持し安定した発話を行えるとともに，シナリオに合わせて演技を行い，映像に合わせてアフレコを行うといった特別な技能を必要とするため，アフレコ経験のある声優を対象話者とし1名の声優に付き2種類の声質でそれぞれのシナリオ音声を収録することとした．声優30名を対象としてシナリオ音声を収録し，計60種類の声のバリエーションを有するシナリオ音声データベースを構築した．主観評価の結果，主観年齢は8〜63歳と幅広い年齢層に対応しているとともに，主観性別についても，男性29名，女性31名とほぼバランスがとれていることを確認した．

4.2.2　システムの評価

試作システムを組み込んだ映像作品“Grand Odyssey”（同時体験人数15名）を体験した一般来場者172名を対象に，アンケート調査を実施した．その結果，選択されたシナリオ音声が自分に合っていたと回答した参加者は35%であった．参加者が録音された自分の音声を聞き慣れていないこと，データベースの規模が十分でなかったことが原因として考えられる．その一方で，76%の参加者がキャラクタに割り当てる声優の声が固定であることに比べ，自分に似ている声優を割り当てられることが面白いと感じていることを確認した．これはエンタテインメントシステムとしての期待を示す結果であり，試作システムがエンタテインメント性をより向上させる付加価値として期待できることが分かった[28]．

4.3　声優音声モーフィング

シナリオ音声データベースに含まれる話者の音声を補間しバリエーションを増やすことにより，利用者の音声に近い音色のシナリオ音声を提供できる可能性がある．そこで，STRAIGHTに基づく音声モーフィングを利用して，シナリオ音声データベースから参加者に近い声質の台詞音声を生成する手法について検討した．

STRAIGHTに基づく音声モーフィングは，STRAIGHTスペクトル，非周期性指標，F0の音声パ

NII-Electronic Library Service

Institute of Electronics, Information, and Communication Engineers

ラメータ，時間軸及び周波数軸の非線形伸縮量を手動で定めた共通のモーフィング率により複数の音声間のモーフィングを実現する．本研究課題では，対象となる参加者の音声に近づくように，台詞音声データベース内の複数話者の台詞音声（同一発話内容）のモーフィング率を自動で決定する新しい手法を実現した．モーフィング率は，対象となる参加者音声とモーフィングによって生成される音声のスペクトル距離が小さくなるように準ニュートン法にて逐次的に求めた．

利用者音声（目標話者音声）と提案手法を用いたモーフィング音声とのスペクトル距離を比較した結果を図18に示す．図から，提案手法を用いることでスペクトル距離が最も近い単一音声よりモーフィング音声のスペクトル距離が小さくなっており目標話者音声により近い音声が生成されることを確認した．また，60種類のシナリオ音声データベースを基に4.1に述べた類似性評価技術を用いて選択した音声と，提案手法により生成されたモーフィング音声とを話者性（どちらの音声が目標話者音声に似ているか）の観点から評価した．その結果（図19），読み上げ音声では69.8%，背景音・効果音（BGM/SE）を重畳したシナリオ音声では56.5%の割合で提案手法により生成されたモーフィング音声の方が似ていると評価され，より利用者の音声に近い音色のシナリオ音声を提供できることを確認した[29]．

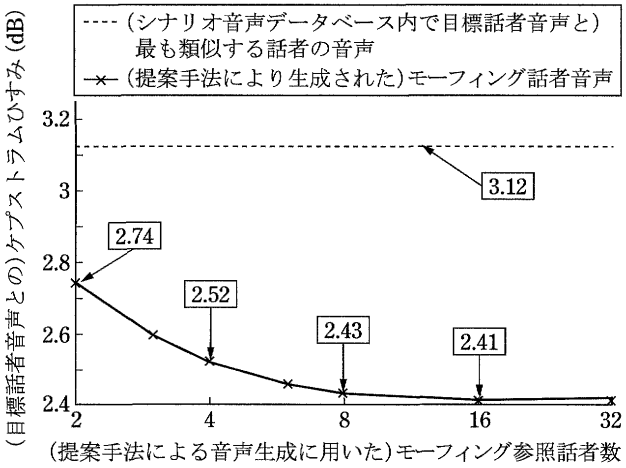


図18　目標話者音声とのケプストラムひずみ

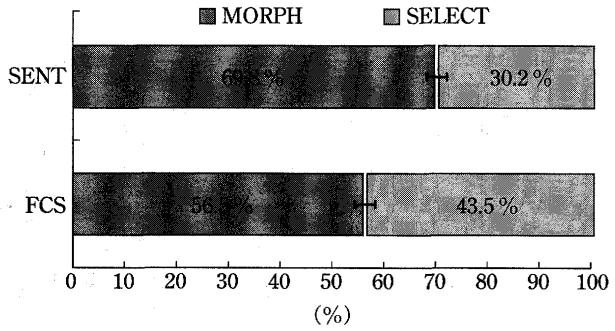
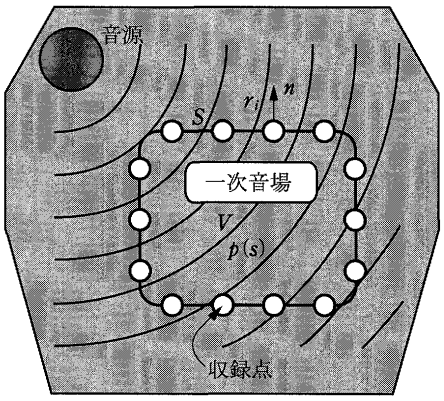


図19　類似話者選択結果音声（SELECT）とモーフィング結果音声（MORPH）との話者性を比較したパフォーマンス評価及び95%信頼区間　SENT：ATR音声データベースBセットから抽出した16文の音声，FCS：シナリオ音声データベースから抽出した39音声サンプルに，劇中のBGM/SEを重畳したもの．

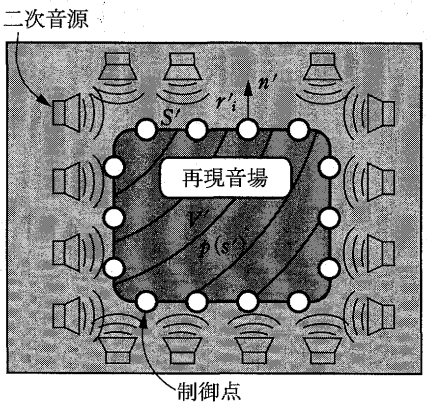
5. ダイブイントゥザムービーのための超臨場感音響処理技術

5.1 境界音場制御原理に基づく音場再現

DVDやBlu-Rayに対応したコンテンツが制作され，自宅などでも5.1 chや7.1 ch，あるいはそれ以上のチャネル数のサラウンドを比較的簡単に体験できるようになってきた．サラウンドは臨場感を高める上で重要な要因であり，体験した人々はあたかも「その場」にいるかのような感覚を得ることができる．しかしながら，現在のサラウンドシステムは実際に「その場」に広がっていた音の波面（音場）を再現しているわけではない．本研究課題では音場を物理的に忠実に記録・再現することを目指し，境界音場制御の原理[30]（BoSC：Boundary Surface Control）に基づきシステムを構築した．物理的に



(a) コンサートホールのような音場



(b) 自宅のリビングルームのような音場

図20　BoSCによる音場再現の概略

NII-Electronic Library Service

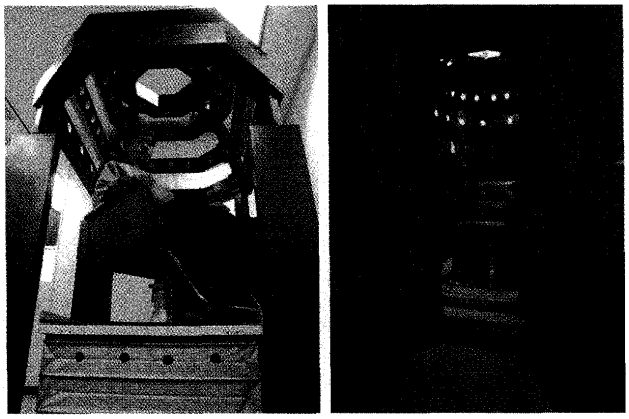
Institute of Electronics, Information, and Communication Engineers

忠実な音場を再現することにより，体験者により深い没入感を与えることができる．

BoSC に基づく音場再現システムの概略図を図 20 に示す．(a)が例えばコンサートホールのような音場であり音源から音波が広がっている．(b)が例えば自宅のリビングルームなどである．このとき，(a)の境界 S で囲まれた領域 V の音場を，(b)の境界 S' で囲まれた領域 V' で再現することがシステムの目的である．従来の波面合成システムは境界 S 上の収録点に対応する S' 上の点にスピーカ（二次音源）を設置して音場再現を行う．しかしながら，原理的には点音源や双極音源を設置する必要があり一般的なスピーカでは正確に波面を合成することはできない．対して BoSC は，境界上にスピーカを設置する代わりに境界の周囲に並べたスピーカから記録した音圧と粒子速度が境界 S' 上の制御点で等しくなるように音を出力することで領域 V' の音場を V と等しくする．スピーカからの出力は二次音源から制御点までの伝達関数行列の逆システムと記録した音圧と粒子速度を畳み込むことにより定まる．ただし，境界 S と境界 S' の形状は合同であり，収録点と制御点の位置も一致しているものとする．また粒子速度は境界面を挟む 2 点の音圧から計算する．

5.2 スピーカアレー，マイクロホンアレーの構築

5.1 で述べた BoSC 型音場再現システムを実現するため，二次音源となるスピーカアレー，収録点及び制御点となるマイクロホンアレーを構築した[31]．構築したスピーカアレーを図 21(a)に示す．スピーカアレーは 4 層のドーム部分と，ドーム部分を支える 4 本の柱からなる．4 層のドーム部分には，それぞれ上から 6 個，16 個，24 個，16 個のフルレンジユニットが設置されている．4 本の柱部分にはそれぞれ 2 個のサブウーファユニットが設置されており，ドーム部分と合わせスピーカ数の合計は 70 個となる．スピーカシステムの外寸は，幅約 1,190 mm，奥行き約 1,600 mm，高さが約 2,010 mm である．ドーム部分と柱部分は内部が空洞になっており，それぞれがエンクロージャの役割を果たす．また，それぞれが分解することが可能であり比較的容易に運搬することもできる．スピーカアレーは試聴時の暗騒音の影響を小さくするため，図 21(b)に示す組立式の簡易型防音室内に設置されている．



(a) 音場再現用スピーカアレー（木製，ドーム型）　(b) 音場再現ルーム

図 21 BoSC 型音場再現システム

構築したマイクロホンアレーを図 22 に示す．受聴者の頭部の周辺の三次元音場を記録し再現するため直径約 46 cm の C80 フラーレンの分子構造を基に下部の 10 個の頂点を取り除いた 70 個の頂点を持つ構造とした．無指向性マイクロホンを各頂点に設置し境界面上では音圧のみを記録し再現する．音圧のみを制御する場合にも境界の形状から定まる固有周波数を除き音場を記録・再現することができる．

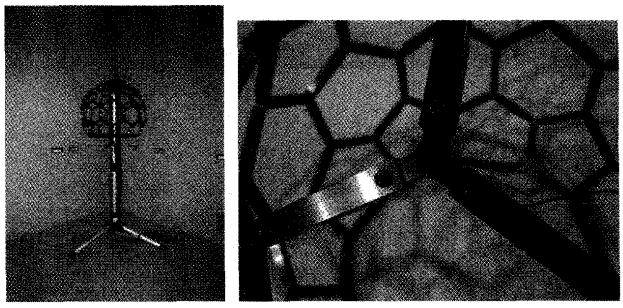
本研究課題では，方向定位実験[31]，相対距離知覚の主観評価実験[32]，再現音場の総合的な印象の主観評価実験[33]，再現音場の音圧分布の測定による物理的な評価実験[34]などを実施し，構築した BoSC 音場再現システムが高い音場再現性能を持つことを確認した．

5.3 三次元音響空間データベースの構築

構築したマイクロホンアレーを用いて三次元音響空間の収録を行った．マイクロホンアレーには DPA 4060-BM が 70 個設置されており，プリアンプは YAMAHA MLA8 を用いた．各チャネルの信号は 3 台の MOTU 24 I/O を用いて同期収録される．サンプリング周波数は 48 kHz，量子化ビット数は 16 bit もしくは 24 bit とした．

・ オーケストラの練習風景の収録

京都府相楽郡精華町にあるコンサートホールにおいて約 70 名の市民オーケストラの練習風景の収録を行った．



(a) マイクロホンアレー外観　(b) マイクロホン拡大図

図 22 C80 フラーレンの分子構造を基に構築したマイクロホンアレー　三次元音場の収録，逆システム設計に使用．

NII-Electronic Library Service

Institute of Electronics, Information, and Communication Engineers

収録は，コンサートホールの客席に加え舞台上の指揮者の正面にマイクロホンを設置して行った．

・ 自然音の収録

(株)国際電気通信基礎技術研究所（ATR）の中庭において自然音の収録を行った．鳥やセミ，かえるの鳴き声，飛行機が空を飛ぶ音，人の声，足音などを収録した．

5.4 システム評価

収録したデータはアウトリーチ活動をはじめ様々な場所でデモンストレーションに使用している．2008年度のアウトリーチ活動では自然音，オーケストラの練習風景（観客席，舞台上）を音源として用いた．来場者に体験後に行ったアンケート結果の一部を以下に示す．

まず「自然の中にいるように感じたか」の質問に対して，①感じた人255人（65%），②少し感じた人111人（29%），③それ以外の人24人（6%）となり，ほぼ全ての体験者が自然の中にいるように感じたと回答していることが分かる．また，印象に残った音という質問に対しても，鳥の鳴き声や人の声などを挙げる回答が数多く見られた．

次に「オーケストラの真ん中にいる感じがしたか」の質問に対して，①感じた人258人（66%），②少し感じた人89人（23%），③それ以外40人（11%）という結果となり，ほぼ90%の体験者がオーケストラの真ん中にいるように感じたと回答していることが分かる．特に今回のオーケストラの配置では，収録位置の後にフルートやオーボエなどの管楽器が位置していたが，そのような音が後方から聞こえる体験に対する驚きや感動の回答が数多くあり，自然の音と同様に多くの体験者が印象に残った音に挙げていた．

6. ダイブインントゥザムービーのための超臨場感映像技術

筆者らは，映画の登場人物の視点から見た周囲360度の映像を記録し，配役を中心とする全方向の映像を呈示できるシステムを開発した．本システムにより，視聴者は登場人物になったつもりで映画の中に飛び込み，高い臨場感の映画鑑賞を楽しむことができる．本章では，超臨場感映像技術を実現するための要素技術を順に述べる．

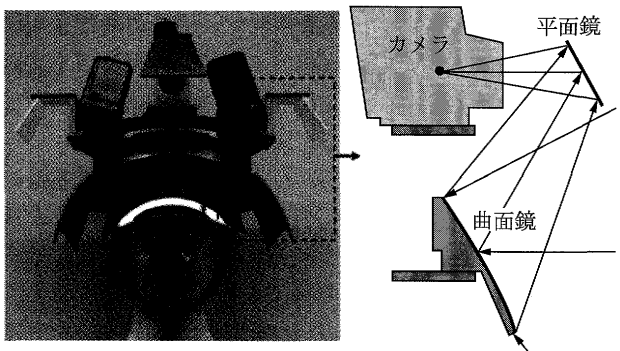
6.1 主観視点全方位映像収録装置

登場人物の主観視点における高品位な全方位映像を撮影するため，以下の性能を有する撮影装置が必要となる．

① 高解像な全方位映像を実時間で撮影できること
② 装着者視点と撮影光学系視点が一致していること
③ 可搬性が高く長時間撮影が簡単に行えること

これらの条件を満たすために，筆者らは図23に示すような，新規に設計した曲面鏡を用いた主観視点全方位映像収録装置 FIrst-Person Perspective Omnidirectional camera : FIPPO を試作した[35],[36]．

試作機はヘルメットの前後左右に4個の光学ユニットを放射状に配置した構成で，各光学ユニットは撮影用のハイビジョンビデオカメラ SANYO Xacti DMX-HD2，及び平面鏡・曲面鏡によって構成されている．曲面鏡はビデオカメラによる画角を広げ，更に空間に対する撮影分解能を均一にするという光学的仕様に従って設計されている[37]．また平面鏡は光学系をコンパクトに収める働きをする．これにより頭部に搭載した4台のカメラで方位によらず安定した画質の全方位映像を記録することができる．反射鏡を用いることでカメラシステムの視点位置を実際の撮影者の視点位置に近づけることが可能となったため，従来の頭頂部搭載カメラなどで生じていた視線のずれが解消されている．また，パソコンなどの追加の記録装置や，有線による外部電源供給なしに単体で長時間撮影が行える．FIPPO の詳細なスペックを表1



(a) 外観　(b) 2枚の鏡による光路

図23 主観視点全方位映像収録装置（FIPPO）

表1 開発した撮影システム（FIPPO）のスペック

項目	内容
総重量	2.5 kg
大きさ	W40 cm × D40 cm × H24 cm
視野角（経度方向）	360度
視野角（緯度方向）	−31.8度〜+26.6度
連続撮影時間	85分
総解像度	約4,480×630pixel
出力形式	MPEG-4video
撮影者視点からの距離	約5.0 cm

NII-Electronic Library Service

Institute of Electronics, Information, and Communication Engineers

図 24　バスケットボールシーンの収録映像の例

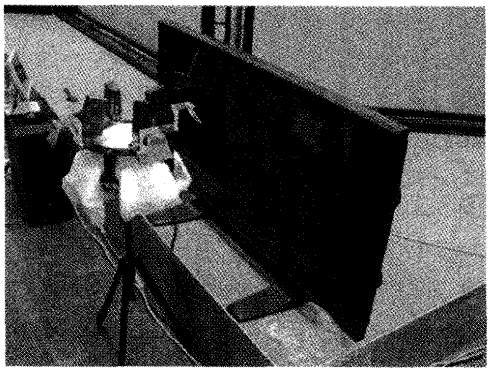
に示す．

6.2　主観視点映像の撮影と編集

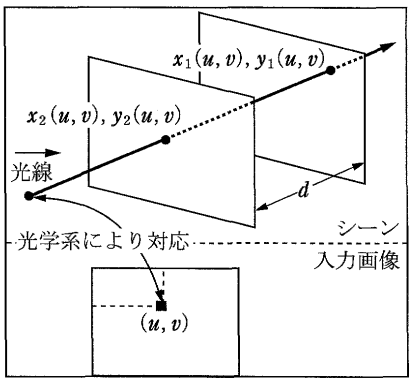
FIPPO を用いて大きく分けて 2 種のコンテンツ映像を収録した．まず日常生活で比較的よく体験するシーンとして，紅葉・水辺・神社参道などのシーンを収録した．一方，普段余り体感することのないシーンとしてスポーツゲームシーンを収録した．一般に，スポーツゲームはフィールドの外側に配置されたカメラの映像を用いて観戦される．本研究では選手や審判に記録装置を装着して収録を行うことで，主観視点の仮想体験を試みた．図 24 は実際の収録映像の例を示している．

曲面鏡を用いて撮影しているため，収録された映像には幾何ひずみが生じている．そこで，事前に画像とシーンの幾何射影を測定しておき，コンテンツ用映像にはその逆変換を適用することで補正を行った．図 25 のように FIPPO の視野に設置されたモニタにパターンを表示し，映像上での画素位置とモニタ上での二次元位置を対応させる．奥行を変化させて 2 回以上測定することで，画像上の画素とシーン中における入射光線の対応を推定した．この対応関係を用いれば，ひずんだ収録映像をパノラマ映像や透視投影映像などに補正することができる．図 26 に幾何補正前後の画像例を示す．シーン中で直線だった部分が正しく直線に戻されていることが確認できる．

また，FIPPO は同一型番のビデオカメラ 4 台で構成されているが，個体差により各々の収録映像は若干色味が異なる．そのため，各ビデオカメラの色空間の間にアフィン変換が成立すると仮定して，図 27 のようにカラーチャートを撮影した映像を用いて色補正を行った．補正前は 2 カメラ映像の境界領域で色が異なるが，補正後は均一になっていることが分かる．



(a) 実験の様子



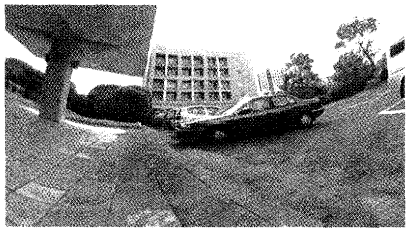
(b) 対応推定手法

図 25　幾何ひずみの補正のための，画素と光線の対応推定

これらの幾何・色補正を施してパノラマ映像を作成した例を図 28 に示す．4 台のビデオカメラで別々に撮影された映像から違和感なく全方位映像を作成できることが確認できる．

NII-Electronic Library Service

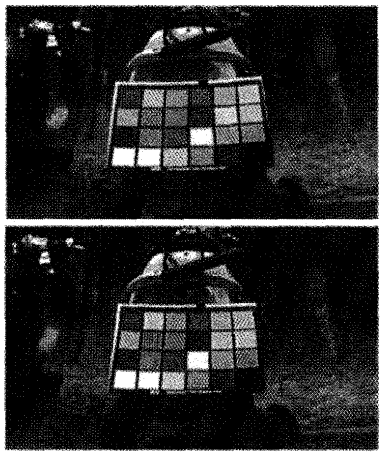
Institute of Electronics, Information, and Communication Engineers



(a) 補正前(入力画像)

(b) 補正後

図 26　透視投影画像への幾何ひずみ補正結果



(a) 撮影したカラーチャート

(b) 色補正前

(c) 色補正後

図 27　カメラ間の色補正

6.3　主観視点映像呈示システム

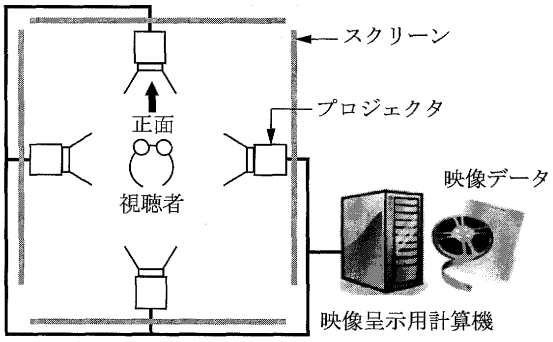
視聴者の前面に設置されたモニタに映像を呈示する方法では，全方位映像が持つ臨場感が損なわれてしまう．そこで視聴者に対して広視野の映像を呈示することで，撮影者主観の高臨場感再現を実現した[38]．

本研究では複数視聴者の同時体験を目的とし，平面スクリーンで四方が囲われるような全方位呈示装置を試作した．図 29 に試作した呈示装置である主観視点全方位シアターのモデル図，及び内部の様子を示す．

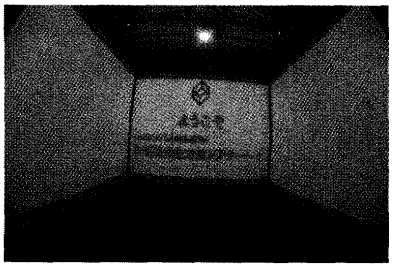
一方，視聴者の周囲に複数台のスピーカを配置することで三次元音場を再現した．スピーカが映像に干渉しないようにスクリーンに音抜け加工を施し，スクリーンの外側かつ視聴者の耳の高さにスピーカを配置した（図 30）．

7. まとめと今後の展望

2009 年 3 月 20～22 日の三日間，プロジェクト終了評



(a) シアターのモデル図



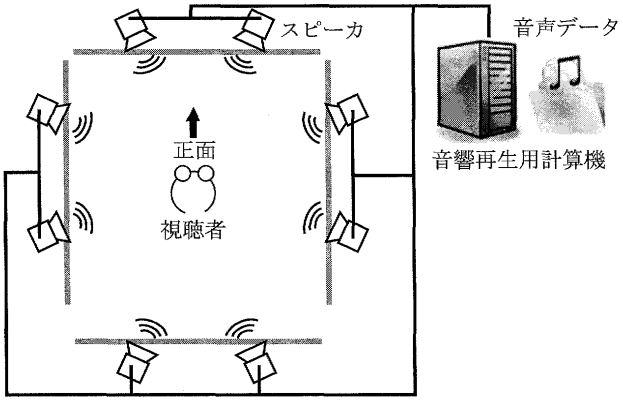
(b) 内部の様子

図 29　主観視点映像呈示システム

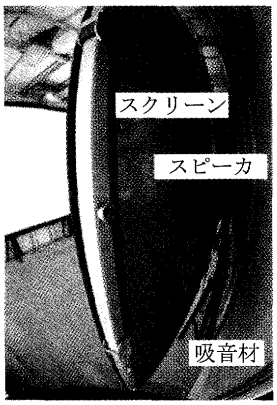
図 28　幾何・色補正によって作成したパノラマ映像

NII-Electronic Library Service

Institute of Electronics, Information, and Communication Engineers



(a) 主観視点型音場再現システム概要



(b) スピーカ設置の様子

図 30　音響再生構成

価実験を日本科学未来館を舞台に実施した．体験者は三日間で 1,610 名に上った．ここで，六つの展示を行った．特にインスタントキャスティングによる映画への没入体験では，参加者の顔形状，表情，肌の質感，歩容，声の特徴に関する個人データを計測し，毎回 10 名の参加者に対して，移動や待ち時間も含む 10 分から 15 分の時間以内で個人 CG キャラクタモデルを全自動生成し，ムービー上映時には，リアルタイムアニメーション生成技術とリアルタイムレンダリング技術によって参加者個人の特徴を表現したキャストとして演技させるシステムを動作させることに成功した．展示の内容は以下のとおりである．

① インスタントキャスティング体験（単純なセンテンスの「音声データ」，歩き方を計測した「歩容データ」，肌の質感を計測した「肌データ」に加え，顔の三次元形状取得データ，表情変化のビデオ画像から即座に分析した結果から個性を反映する CG キャラクタモデルを生成する．本編上映時にはリアルタイムレンダリング・キャラクタ動作を生成して，映画上映を実施した．）

② アフレコ体験（短いセンテンスの音声収録．プロの声優にしゃべり方を変換する音声モーフィング実験）

③ 主観視点全方位シアター（全周囲のスクリーンで人物の視点から見た映像と音響を再現）

④ 歩き方の個性計測（歩き方の特徴を解析）

⑤ 三次元音場再現システム（音の空間的な広がりを再生）

⑥ 肌の質感計測（肌の表面下散乱特性を計測，拡散反射成分と鏡面反射成分とを分離）

自己を映像シーンの中で発見することができたかという評価実験の結果，2006 年度に実施した「フューチャーキャスト」評価結果が 57% であったのに対して，2008 年の改良バージョンで 87% への性能向上を達成した．他に，体験してとても楽しかった（34%），他の作品で楽しんでみたい（20%），アニメに応用したら面白そう（14%），友達の方が分かりやすい（10%），大いに感動した（5%），元気が出た（7%），などの感想が寄せられた．

映画キャスティングのみならず，様々なアプリケーションへの本技術の適用に関して，現在，検討を進めている．本技術が新しいエンタテインメント領域の開拓，イノベーションをもたらすことを期待して，今後も研究を継続してゆく予定である．

文　　献

(1) L. Tao, S. Morishima, and A. Maejima, "The effects of virtualcharacters on audiences' movie experience," Interacting with Computers, vol. 22, no. 3, pp. 218-229, 2010.

(2) A. Maejima, S. Wemler, T. Machida, M. Takebayashi, and S. Morishima, "Instant casting movie theater : The future cast system," IEICE Trans. Inf. & Syst., vol. E91-D, no. 4, pp. 1135-1148, April 2008.

(3) 前島謙宣，森島繁生，"最適化局所アフィン変換に基づく正面顔レンジスキャンデータからの頭部モデル自動生成，" 画像電子学会論文誌，vol. 38, no. 4, pp 404-413, 2009.

(4) C. Zhu, R.H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778 : L-BFGS-B : Fortran subroutines for large-scale bound-constrained optimization," ACM Trans. Math. Softw., vol. 23, no. 4, pp. 550-560, 1997.

(5) B. Allen, B. Curless, and Z. Popovic, "The space of human body shapes : reconstruction and parameterization from range scans," ACM SIGGRAPH 2003, pp. 587-594, 2003.

(6) 前島謙宣，森島繁生，"顔形状事前知識に基づく顔画像からの 3 次元顔モデル高速自動生成，" 第 16 回画像センシングシンポジウム SSII2010，no. IS3-13, 2010.

(7) 久保尋之，石橋　康，前島謙宣，森島繁生，"表情筋運動モデルの過渡特性を考慮した表情アニメーション"，VC/GCAD 合同シンポジウム，no. 20, June 2008.

(8) 久保尋之，鈴木孝章，石橋　康，森島繁生，" 表情筋モデルを用いた動画像からの表情合成，" 画像電子学会　VC ワークショップ，vol. 37, no. 6, p. 1061, 2008.

(9) J.Y. Bouguet, Pyramidal implementation of the Lucas Kanade feature tracker description of the algorithm, Intel Corporation, Microsoft Research Labs, 1999, Open CV Documents.

(10) A. Maejima, T. Suzuki, H. Yarimizu, H. Kubo, and S. Morishima," An automatic individual character head modeling for instant casting," ACCV Workshop, no. IW1-2, 2009.

(11) H. Yarimizu, Y. Ishibashi, H. Kubo, A. Maejima, and S. Morishima,

NII-Electronic Library Service

Institute of Electronics, Information, and Communication Engineers

"Muscle-based facial animation considering fat-layer structure captured by MRI," Proc. SIGGRAPH Posters, no. 19A, 147-yarimizu. pdf, Aug. 2009.
(12) 馬場葉子，間下以大，向川康博，八木康史，"大規模データベースを用いた肌の反射・散乱光の統計的解析，" 画像の認識・理解シンポジウム（MIRU2009），no. IS2-46, pp. 1164-1171, 2009.
(13) S.K. Nayar, G. Krishnan, M.D. Grossberg, and R. Raskar, "Fast separation of direct and global components of a scene using high frequency illumination," ACM SIGGRAPH2006, pp. 935-944, 2006.
(14) 間下以大，向川康博，八木康史，"多層表面下散乱モデルによる皮膚の異方性散乱と不均一性の表現，" 画像の認識・理解シンポジウム（MIRU2009），no. IS2-53, pp. 1216-1222, 2009.
(15) 馬場葉子，向川康博，八木康史，"化粧と肌の2層構造からなる化粧肌反射特性モデル，" MIRU 2010, no. IS3-19, pp. 1679-1685, 2010.
(16) 久保尋之，土橋宜典，森島繁生，"曲率に依存する反射関数を用いた半透明物体の高速レンダリング，" 信学論(A)，vol. J93-A, no. 11, pp. 708-717, Nov. 2010.
(17) D. Gafurov, K. Helkala, and T. Sondrol, "Biometric gait authentication using accelerometer sensor," J. Comput., vol. 1, no. 7, pp. 51-59, 2006.
(18) A. Bobick, A. Johnson, "Gait recognition using static activity-specific parameters," Proc. IEEE Conf. on CVPR, vol. 1, pp. 423-430, 2001.
(19) 奥村麻由，槇原　靖，八木康史，中村慎介，森島繁生，"視聴者参加型デジタルエンタテインメントのためのオンライン歩容個性計測，" MIRU2009, no. IS2-70, pp. 1336-1343, July 2009.
(20) 槇原　靖，佐川立昌，向川康博，越後富夫，八木康史，"周波数領域における方向変換モデルを用いた歩容認証，" 情処学論コンピュータビジョンとイメージメディア，vol. 48, No. SIG 1 (CVIM 17)，pp. 78-87, Feb. 2007.
(21) 万波秀年，槇原　靖，八木康史，"歩容における性別・年齢の分類と特徴解析，" 信学論(D)，vol. J92-D, no. 8, pp. 1373-1382, Aug. 2009.
(22) OU-ISIR Gait Database, "http://www.am.sanken.osaka-u.ac.jp/GaitDB/index.html"
(23) 中村槙介，森島繁生，"歩行における知覚的類似性尺度に基づく個人性を強調した動作合成手法，" 画電学誌，vol. 39, no. 5, pp. 615-620, 2010.
(24) L. Kovar and M. Gleicher, "Flexible automatic motion blending with registration curves," SIGGRAPH2003, pp. 214-224, Aug. 2003.
(25) 河原英紀，"Vocoderのもう一つの可能性を探る—音声分析変換合成システムSTRAIGHTの背景と展開—，" 音響誌，vol. 63, no. 8, pp. 442-449, 2007.
(26) H. Kawahara, I. Masuda-Kastuse, and A. de Cheveigné, "Reconstructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction," Speech Commun., vol. 27, no. 3-4, pp. 187-207, 1999.
(27) 中村　哲，足立吉弘，川本真一，森島繁生，"個人特徴を反映するマルチメディアコンテンツ制作のための話者類似性評価技術，" 第17回信学バイオメトリックシステムセキュリティ研資，pp. 33-36, 2009.
(28) 足立吉広，大谷大和，川本真一，四倉達夫，森島繁生，中村哲，"個人の音声を反映する映像エンタテインメントシステム，" 情処学論，vol. 49, no. 12, pp. 3908-3917, 2008.
(29) 川本真一，足立吉広，大谷大和，四倉達夫，森島繁生，中村哲，"来場者の声の特徴を反映する映像エンタテインメントシステムのための台詞音声生成システム，" 情処学論，vol. 51, no. 2, pp. 250-264, 2010.
(30) 伊勢史郎，"キルヒホッフ-ヘルムホルツ積分方程式と逆システム理論に基づく音場制御の原理，" 音響誌，vol. 53, no. 9, pp 706-713, 1997.
(31) S. Enomoto, Y. Ikeda, S. Ise, and S. Nakamura, "Three-dimensional sound field reproduction and recording systems based on the boundary surface control principle," Proc. of 14th Int. Conf. Auditory Display, pp. o_16, Paris, 2008.
(32) 榎本成悟，池田雄介，伊勢史郎，中村　哲，"境界音場制御の原理を用いた音場再現システムにおける距離感の再現精度に関する評価，" 音響論集，pp. 725-726, Sept. 2008.
(33) 池田雄介，榎本成悟，伊勢史郎，中村　哲，"多チャンネル三次元音場再生システムの再生方式の違いによる再現性能評価，" 信学技報，EA2008-114, pp. 77-82, Dec. 2008.
(34) 榎本成悟，池田雄介，伊勢史郎，中村　哲，"境界音場制御の原理を用いた音場再現システムにおける再現領域の物理的な評価，" 音響論集，no. 3-P-9, pp. 855-856, March 2008.
(35) 近藤一晃，向川康博，八木康史，"主観視点全方位映像撮影のための装着型システムの開発，" 画像の認識・理解シンポジウム（MIRU2008），no. IS5-43, pp. 1656-1661, 2008.
(36) K. Kondo, Y. Mukaigawa, and Y. Yagi, "Wearable imaging system for capturing omnidirectional movies from a first-person perspective," Proc. VRST2009, 2009.
(37) 近藤一晃，向川康博，八木康史，"表面勾配の可積分性を考慮した自由形状ミラー設計法，" MIRU2008, no. OS2-2, pp. 42-49, 2008.
(38) 近藤一晃，向川康博，池田雄介，榎本成悟，伊勢史郎，中村哲，八木康史，"主観視点全方位映像と三次元音場による高臨場感シーン再現，" 画像の認識・理解シンポジウム(MIRU2009)，no. IS3-52, pp. 1729-1736, 2009.

（平成22年10月3日受付　平成22年11月19日最終受付）

森島　繁生（正員）
1987東大大学院工学研究科博士課程了．工博．同年成蹊大専任講師，1988同助教授，2001同教授．2004早大教授，1994～1995トロント大客員教授．現在，明大非常勤講師，新潟大非常勤講師．CG, CV, HIの研究に従事．1991年度本会業績賞，2010テレコムシステム技術賞各受賞．

八木　康史（正員）
1985阪大大学院基礎工学研究科了．三菱電機，阪大・基礎工を経て，2003から同大学産業科学研究所教授．コンピュータビジョン研究に従事．1996年度本会論文賞，MIRU（2008長尾賞，2010優秀論文賞），情報処理学会フェロー．博士（工学）．

中村　哲（正員）
1981京都工繊大卒．1992京大博士（工学）．シャープ中央研究所，奈良先端大，ATRを経て，現在独立行政法人情報通信研究機構けいはんな研究所長，知識創成コミュニケーション研究センター長．ATRフェロー．音声言語情報処理の研究に従事．2010文部科学大臣賞受賞．

伊勢　史郎
1984早大・理工・電子通信卒，～1986コルグ株式会社，～1988早大大学院理工学研究科修士課程電気工学専攻，～1991東大大学院博士課程建築学専攻，～1994早大・理工・客員講師，～1998奈良先端大助手．1998～京大大学院工学研究科建築学専攻准教授．工博．

向川　康博（正員）
1997筑波大大学院工学研究科博士課程了．同年～2002岡山大助手，2003～2004-10筑波大講師．阪大助教授を経て，2007-04から同大学准教授．博士（工学）．情報処理学会，日本バーチャルリアリティ学会，IEEE各会員

槇原　靖
2001阪大・工卒．2002，2005同大学院工学研究科博士前期，後期課程をそれぞれ了．博士（工学）．2005，2006，2007にそれぞれ同大学産業科学研究所特任助手，助手，助教となり現在に至る．歩容認証に関する研究に従事．

NII-Electronic Library Service

Institute of Electronics, Information, and Communication Engineers

間下　以大（正員）

2006 阪大大学院博士後期課程了．博士（工学）．同年同大学特任研究員．2008 同大学助教，現在に至る．コンピュータビジョン，パターン認識に関する研究に従事．ヒューマンインタフェース学会，日本ロボット学会各会員．

近藤　一晃（正員）

2007 阪大大学院情報科学研究科博士後期課程了．同年同大学産業科学研究所特任研究員．2009 京大学術情報メディアセンター助教就任後現在に至る．反射屈折光学系，知的行動支援に関する研究に従事．博士（情報科学）．情報処理学会会員．

榎本　成悟（正員）

2005 京大大学院工学研究科博士後期課程了．博士（工学）．2005（株）国際電気通信基礎技術研究所研究員，2009 独立行政法人情報通信研究機構専攻研究員，現在に至る．日本音響学会粟屋学術奨励賞受賞．日本音響学会，米国音響学会各会員．

川本　真一（正員）

2005 北陸先端大大学院博士後期課程了．博士（情報科学）．（株）国際電気通信基礎技術研究所，独立行政法人情報通信研究機構を経て，2010 から北陸先端大助教．音声・マルチモーダル情報処理の研究に従事．情報処理学会，日本音響学会各会員．

四倉　達夫（正員）

1998 成蹊大・工・電気電子卒．2000 同大学院修士課程了．2003 同大学院博士課程了．博士（工学）．2003（株）国際電気通信基礎技術研究所音声言語コミュニケーション研究所．2009 から（株）OLM デジタル研究開発部門．

池田　雄介

2001 早大・理工・情報卒．2003 同大学院修士課程了．2007 同大学院博士課程了．博士（国際情報通信学）．2007（株）ATR 研究員．2009 から独立行政法人情報通信研究機構専攻研究員．音場再現，音場可視化に関する研究に従事．日本音響学会会員．

前島　謙宣（正員）

2010 早大大学院理工学研究科博士後期課程了．同大学・理工学術院・助手を経て，現在，同大学・IT 研究機構・次席研究員．博士（工学）．CG による人物表現に関する研究に従事．画像電子学会，ACM 各会員．

久保　尋之（学生員）

2006 早大・理工卒．2008 同大学院博士前期課程了．現在，同大学院博士後期課程に在学中．2008 から日本学術振興会特別研究員（DC1）．

NII-Electronic Library Service