2008－CVIM－161 (19)
2008／1／17

ストーリへの没入感を実現する ダイブイントゥザムービープロジェクト

森島繁生* 八木康史† 中村 哲**
*早稲田大学理工学術院 〒169-8555 東京都新宿区大久保 3-4-1
†大阪大学産業科学研究所 〒567-0047 大阪府茨木市美穂ケ丘 8-1
** ATR SLC 〒619-0288 京都府相楽郡精華町光台 2-2-2
E-mail: *shigeo@waseda.jp †yagi@am.sanken.osaka-u.ac.jp **satoshi.nakamura@atr.jp

キーワード 次世代エンタテインメント、個人表情合成、個人動作合成、個人音声合成、頭髪運動合成、３次元音場、没入空間、インスタントキャスティング

アブストラクト 『ダイブイントゥザムービー』は、視聴者自らが映画の登場人物として演技し、それを観客として家族や友人と鑑賞し感動を共有する１つ形態と、忠実な環境を仮想再現することによって観客自らが主人公の存在する空間や音場を体験共有する形態という、２つのスタイルでストーリへの没入感を実現する、世界に類を見ないエンタテインメントの提案である。この実現のためには、視聴者そっくりのキャラクタモデルを短時間のうちに生成し、本人の特徴を忠実に再現しながら実時間合成する必要がある技術開発と、演者を取り巻く映像および音場環境を忠実に記録し、自動的に再現する技術が必要となる。

"Dive into the Move" Project to realize an immersive experience in the story

Shigeo MORISHIMA* Yasushi YAGI† Shinichi KAWAMOTO**
*Waseda University, 3-4-1 Okubo Shinjuku, Tokyo 169-8555 Japan
†The Institute of Scientific and Industrial Research, 8-1 Mihogaoka, Ibaraki, Osaka, 567-0047
** ATR SLC., 2 Hikaridai, Seikacho, Sorakugun, Kyoto, 619-0288
E-mail: *shigeo@waseda.jp †yagi@am.sanken.osaka-u.ac.jp **shinichi.kawamoto@atr.jp

Keyword Next Generation Entertainment, Personal Expression Synthesis, Personal Gesture Synthesis, Personal Voice Synthesis, Hair Motion Synthesis, 3D Acoustic Field, Immersive Space, Instant Casting

Abstract "Dive into the Movie" is world new entertainment system which can provide two styles of immersion experience into the story by giving a chance to audience to share an impression with his family or friend by watching a movie in which all audience can participate in the story as movie casts and by giving a chance to experience the panoramic scene and acoustic environment surrounding with the movie cast by special environment capturing and playback system. To realize this system, several techniques not only to model and capture the personal characteristics instantly in face, body, gesture and voice, but also to capture and playback high fidelity acoustic and visual environment automatically.

1. はじめに

映画の登場人物に扮して、宇宙を駆け巡る体験をしてみたい。あの有名な俳優と映画の中で共演してみたい。そのような希望は誰しもが一度は抱いたことがあると思う。本稿で述べるダイブイントゥザムービーはそのような希望を叶えてくれる世界に類を見ないエンタテインメントシステムの提案である。

このシステムの1つの具体的な実現例は、著者らのチームが中心となって 2005 年に具現化した愛知万博、三井・東芝館のフューチャーキャストシステムである。シアターの定員である 240 名全員が映画に出演できるアトラクションシステムとして具現化され、期間中に 163 万人の観客がこれを体験し、人気のパビリオンの1つとなった。

本システムは、実際に来場者自身を映像の中に登場させ、CG 化された自分の分身が映像中で出演者として演技し、それを家族や友人とともに観客として鑑賞することで、ストーリへの没入感を増大させている点に特徴があった。

また、上映開始までの約 10 分間のプレショー時間内に、来場者の全員の顔を CG モデル化し映像を作り上げなければならないため、顔の正面画像・立体形状の入力から CG 顔モデルの生成、映像の上映までの一連のプロセスを全自動で実行していた点も特徴である。

来場者はまず、少人数のグループに分かれ、プレショールームに誘導される。各プレショールームには 3 次元スキャナが配置されており、1 台のスキャナで 4 人分の顔の 3 次元スキャニングを行い、来場者全員の正面顔画像と顔の立体形状が計測される。

次に、正面顔画像から画像処理を施すことにより、顔の特徴的な部位(目、眉、鼻、口、顔輪郭)の特徴点を抽出し、同時にストーリの配役を決定する際に必要な情報である年齢・性別情報の推定を行う。その後、抽出された特徴点に対して標準的な顔ワイヤフレームモデルを自動整合し、このモデルにテクスチャマッピングを施すことで個人の CG 顔モデルが生成される。

生成された顔モデルを実時間変形し、表情合成とストーリ映像と同じ環境下での陰影付けを施して、ストーリ映像とフレーム単位で同期・融合させることで、来場者そっくりの登場人物が、台詞を喋り、時には感情をあらわにしてストーリの中で演技するシーンを、観客の一人として目の当たりにすることができる。

ストーリ映像の上映には、当然信頼性の高いマシンが要求される。レンダリングクラスタシステムを用いて、システムダウンの無い、安定した上映が可能となった。6ヶ月間の愛・地球博開催中に 163 万人の観客が映画の登場人物となる体験をしたがシステム障害は一度も報告されなかった。

またシステムの性能を検証するために、CG 化された来場者の顔モデルが上映に耐えうるか検証したところ、平均 93.5%の確率でシアターに出演が可能であるという結果が得られた。

本システムは、実時間処理によってモデル合成とレンダリング処理を行うゲーム的な要素と大勢の観衆に演技を披露できる演劇的な要素、そして感動をよぶストーリを大勢の観客に映像として同時に提示できる映画の要素、これら全ての特徴を兼ね備えた全く新しいエンタテインメントシステムとして具現化された。

2006 年 3 月よりこの同じシステムが長崎県佐世保市のハウステンボスにて常設展示されている。ここではレンジスキャナのシステムを見直し、スキャンからモデル完成までの時間が大幅に短縮されたことから、万が一の場合の再撮影することが可能となり、シアター出演確率は 100%を実現している。

2. ダイブイントゥザムービープロジェクト

このような背景で、「新映像技術ダイブイントゥザムービーの研究」プロジェクトは、文部科学省科学技術振興調整費プログラムとして、2006 年度後期より開始された。図 1 に、プロジェクトの概要を示す。

このプロジェクトでは、フューチャーキャストシステムをさらに発展させることを目的とし、2 つの課題の実現を目標としている。

1つは、視聴者参加型の映像体験をさらに発展させ、登場人物となる参加者自身の個性をより強く反映させ、自分自身がストーリのなかに役者として参加しているという意識をより強く持たせることを実現とする。さらにもう1つは、役者や映像カメラマンの立ち居地や収録環境をそのまま忠実に再現する3次元仮想没入空間の実現である。

前者では、顔の形状のみならず、顔表情表出の特徴、全身の形状、動作の特徴、声の特徴などに個性を反映させるための要素技術開発を含む。ただし、極めて短時間にかつ被験者に負担をかけず



図１　ダイブイントゥザムービープロジェクトの概要

に特徴抽出を実現しなければならない。一方、後者では役者の目線から見た周囲環境および音場の忠実な収録・再現技術の開発が必要となる。

3. フューチャーキャストの映像技術

フューチャーキャストシステムは、視聴者参加型のエンタテインメントシステムであるが、大人数の参加を実現可能とするためにオペレータの介入を全く必要とせず、全ての処理が全自動で実行される点が特徴である。

3.1. 顔の3次元スキャニング

半円上に配置された７台のデジタルカメラ（愛・地球博版）もしくは１台のカメラ（ハウステンボス版）から撮影された縞模様パターン投影画像の処理によって顔の３次元計測を実施する。正面顔画像のフラッシュ撮影データに基き１画素ごとに奥行きデータとの対応が取られる。この各画素位置に配置された奥行きのデータおよび色のデータは、次の個人ＣＧモデル生成プロセスに受け渡される。

3.2. 個人 CG モデル生成

奥行きデータから個人顔モデルを生成するために、標準となる顔ワイヤフレームを正面顔画像に整合する。この際、目、眉、鼻、口、顔輪郭などの特徴点が自動計算される。奥行き情報は対応する画素位置から決定される。特徴点以外の顔ワイヤフレームの格子点については、BRDF により位置を決定する。目や眉、口の輪郭線上に関しては、わずか数 pixel の整合誤りでも CG 顔モデルにしたときの印象が劣化する。このため、特徴点間をスプライン曲線近似し、標準顔ワイヤフレームモデルとの対応点を決定する。正面顔画像に対して顔ワイヤフレームを自動整合した結果を図２に示す。

これと同時に、正面画像の処理によって、男女判定および年齢推定が行われる。また虹彩の色を自動推定し、眼球のテクスチャを選択する。

立体形状計算用に一人の処理に１台のＰＣを割り当てることで、愛・地球博版では個人の CG ポリゴンモデル生成までに約３分を要した。これが、ハウステンボス版では、アルゴリズム見直しとハードの進化により、約 20 秒までに短縮された。

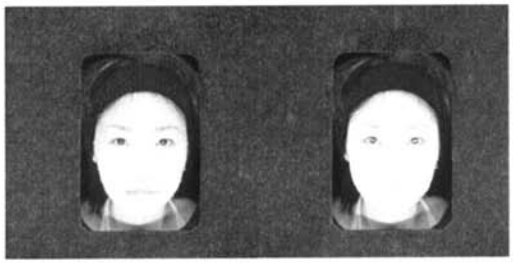


図２　特徴点と標準モデル整合の結果

3.3. ブレンドシェープによる表情合成

怒りや悲しみの表情、発話時の口形状など任意の表情の表現を行うために、基本となる表情パターンを予め基本形状（ターゲットシェープ）として準備しておき、この基本形状のブレンドによって目的となる表情を表現している。グランオデッセイの作品中では 34 種類のターゲットシェープを用意した（図３）。これらの形状は顔の標準ワイヤフレームの格子点の移動量を標準値として定義されている。来場者によって顔のアスペクト比を算出し、各格子点の移動量を個人適応して用いる。図４に表情合成の一例を示す。

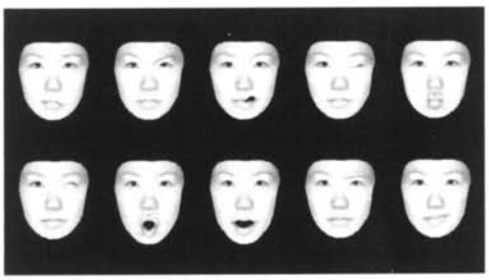


図３　ターゲットシェープの一例



図４　表情合成シーンの例

口の形状と共に、口内部の歯の位置も制御される。今回は、ストーリは予め固定されているので、任意のシナリオに対応する必要はないが、インタラクティブなアプリケーションにも対応可能なシステム構成になっている。

2005 年「愛・地球博」では、「フューチャーキャストシステム」は参加者の顔のみを取り込み、映像として合成する技術であった。

3.4. リアルタイム表情合成

今回の作品 Grand Odyssey というストーリに同期して、各シーン毎の照明環境をはじめとするリアルタイムレンダリングに必要な環境パラメータは時系列の環境シナリオデータとして、予め映像作品作成時にクリエータによって保存されている。また各映像フレームで、セリフに基づく口の動きのパラメータおよび表情のパラメータも時系列のキャストシナリオデータとして保存されている。この口形状パラメータおよび表情パラメータは、先のターゲットシェープのブレンド率によって定義している。よってメインストーリーが開始されると、このレンダリングのための環境シナリオデータと、基本表情ブレンド率および顔の位置と向きの時系列であるキャストシナリオデータに基づいて、毎フレーム同期して、表情変形およびレンダリングが実時間で実行される。図５に映像生成のプロセスを示す。左がプリレンダリングされた背景映像で、実時間生成された顔画像とマスク画像の積により輪郭が切り取られ、コンポジットされて出力画像となる。

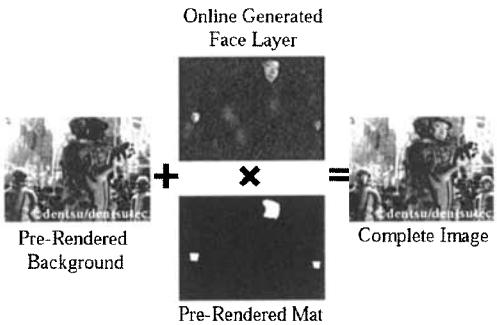


図５　映像生成プロセス

3.5. レンダリングクラスタシステム

三井・東芝館では、合計 12 台の各プロジェクタにつき、スクリーンへの映像送出用 PC が 1 台、

映像生成用 PC が 8 台で構成されている。また全てのスクリーンの映像は同期している。各 PC はギガビットネットワークで接続されている。

映像生成 PC は、Pentiumn4 3.2 GHz、GeForce5950 Ultra、512MB RAM で、ごく一般的な PC で構成されている。上映される画像は、各画像生成 PC により生成され、プロジェクション PC によりタイムコードに同期して投影される。1 フレームの画像生成は 1 台の PC がレンダリングから合成までを担当する。来場者毎に変更される顔モデルは、上映前にプロジェクション PC よりファイルサーバから収集され、各映像生成 PC へ転送される。

本システムでは、限られた期間でのシステムの構築と映像製作を同時進行で進める必要があったため、システムへのパフォーマンス要求は設計時には未知数であり、また設計時には、実装時のマシンスペックを見積もることは困難であった。さらに博覧会の開催期間中はシステムダウンが許されないため、より信頼性のある上映システムを構築する必要があった。そこでスケーラビリティのあるクラスタシステムを採用した。また、レンダリング機能とプロジェクタへの送出機能とを分離することで、システムの信頼性を向上させることが可能となった。レンダリング用 PC は冗長性をもたせ、8 台のうち 5 台の PC が動作していれば上映を続けることが可能であった。

3.6. システム評価

自動生成された来場者の CG 顔モデルが上映に耐えうるかどうか、表情合成をしたときに明らかな顔形状の破綻が見られないかどうか、上映前に必ずアテンダントによる目視確認作業と登場の可否の判断が行われる。登場が認められない場合には、デフォルトデータに差し替えられる。

2005 年 9 月 1 日から愛・地球博の最終日である 9 月 25 日までの、日毎のアテンダント確認結果を集計し、出演確率と 25 日間の平均の出演確率を求めた。評価対象は、25 日間 902 回の上映分となる 21 万 6480 人分の CG 顔モデルである。結果を図 6 に示す。図中横軸は上映日を表しており、縦軸は出演確率を表している。25 日間で平均 93.5%、最高 94.5%、最低でも 92.5%の確率で来場者の CG 顔モデルを映画の中に登場させることができるという結果が得られた。ただし、この確率は、本人が映像中の自分自身を認識できるという意味とは異なっている。

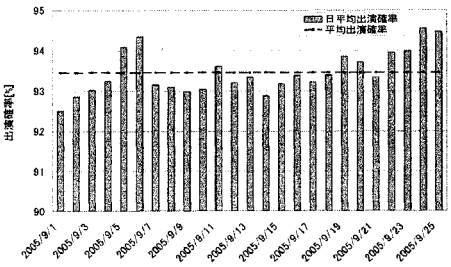


図 6　9 月入場者の出演確率

4. ダイブイントゥザムービーの課題

先のフューチャーキャストシステムでは、スタティックモデルとしての 3 次元立体形状において個人性は忠実に再現されていたものの、キャラクタアニメーション時に参加者がストーリ上の自分自身に気づかない場合が多々あった。また、ヘルメットに包まれた顔のみを実時間合成していたため、頭髪、眼鏡、衣服などの個人の特徴を含んでいなかった。さらに身体サイズやジェスチャに関しては、プリレンダリングによって、キャスト毎に予め定められた動きを毎回の上映ごとに繰り返し、合成された顔の部分のみを個人に応じて変化させていた。これはハードウェアの制約によるものであり、2005 年当時としては限界に近い状況にあった。しかし、GPU 等の進化により、全身のしかも頭髪運動を含むリアルタイムキャラクタ合成も夢物語ではなくなりつつある。

4.1. 観客の声アンケート

2006 年 11 月 25 日、26 日、27 日の 3 日間、日本科学未来館（東京都江東区青海 2-41）にて、サイエンスアゴラ 2006 の一環として、アウトリーチ活動を実施した。ここでは、セミナーとフューチャーキャストシステムの体験会を行い、会場への来場者は、愛・地球博終了以来初めて映画作品『Grand Odyssey』を体験することができた。事前に了承を頂いた来場者対象に、レンジスキャナで顔をスキャニングし、自分の顔がはめ込まれている映画を鑑賞し、万博では不可能であった来場者へのアンケートを実施することができた。

今回、２つのスクリーンを並列・同期して上映し、左には来場者の顔を入れ替えた画像、右側には入れ替えを行わないデフォルトのキャストの画像を上映した。これにより “誰でも演技できる” ことが、より体感できたと思われる。体験者は、

555 名（男性 295（53.5%）、女性 256（46.5%）、無回答 4（0.73%））にのぼり、当日実施したアンケートから以下の考察が得られた。

1）自分自身を識別する特徴として，体格や動作は他の特徴より注目度が若干低くなっている．これは，普段から鏡で確認する機会の多い顔などと比べて，自分自身の体格や動作といった特徴を意識する機会が少ないためであると考えられる．

2）一方で，他者を識別する場合の注目度を自分自身の場合と比べると，動作が 0.01 点，体格が 0.11 点上昇している．特に体格についての上昇率は全特徴の中で最大であり，来場者が映像中から知人を探す際に役立つ特徴であると言える．

3）顔表情は、映像中の自分・他人を見分ける特徴として、顔の次に注目されている。このため、来場者の顔表情に近い表情が合成できれば、自分や知人をより識別することが容易になると考えられる。

4）自分、もしくは他人を見分ける特徴として、音声は顔と表情の次に注目されている。これは、髪型や体格などよりも高い評価であり、映像内で個人を識別する特徴として音声の重要性が改めて確認された。

5）映像中の自分・他人を見分ける特徴としての声は、比較的注目される傾向にある。特に、他人を探す際にはより注目されている。これは少なくとも、来場者に近い声を再生することで、来場者が自分・知人を映像内で探す支援になり、来場者に合わせた台詞音声配役システムの必要性を示す結果といえる。

6）髪型は、声の次に自分・他人の識別のための特徴として注目されている。動作や体型と比較して、日常的に自分が鏡などを利用して眺める機会が多いため、このような結果になったと推測できる。このため出来るだけ個人の趣向を反映した頭髪スタイルの特徴を持つキャラクタを作り上げる必要があると言える。

また、特に顔を撮影した人への質問の中でキャラクタの話す声に関して、各世代合計で 5 割以上、特に20代では7割近くの来場者が違和感を感じていたことがわかった。これは、声の配役に関しては、更なる改善の余地があることを示す。また、違和感を感じる人のうち、7割の人が話し方などではなく声そのものに違和感を感じていることから、発話スタイルよりもまずは来場者に近い声質の台詞音声を再生することが優先課題であることがわかる。

また、来場者からのコメントとしては、以下の2つに大きく絞られた。

・性別の一致を望む声や、大人と子供の違いが分からないといった年齢的な特徴の一致を望む意見があった。特に、声の性別の不一致には非常に敏感であり、「声の性別は一致させてほしい」といった意見が多かった。

・自分や家族・知人・友人の顔を捜すのが精一杯で、ストーリまで見る余裕が無いといった意見も多かった。これより現状のシステムでは「自分・知人を探すゲーム性」の面白さが強く、映画のようにストーリを楽しむエンターテイメント性を伝えるには、技術的な改良が必要であることがわかっている。

4.2. プロジェクトの課題.

上記のアンケート結果を踏まえ、本研究プロジェクトでは、より個人の特徴を強く反映するキャラクタアニメーションの生成手法について検討を行っている。具体的には、まず、リアルタイムレンダリング部分を全身に拡張し、インタラクティブにジェスチャを制御可能とする。次に、表情合成方式を見直し、個人の表情表出特徴を反映するだけではなく、肌の質感等に関しても改良を施す。また、髪型の特徴を CG 再現し物理シミュレーションによって動作に応じた頭髪の運動をリアルタイム生成する。さらに台詞音声を実際の来場者の特徴を反映したものに置き換える。

眼鏡や化粧などの付加、さらにはレンダリング方法に関しても個人の趣向を反映した演出ができる枠組みを生成する。

この研究成果によって、映画の登場人物は、参加者の特徴を有すると同時に、参加者の意図したとおりの演出が可能なオリジナルのキャラクタとなる。このテーマは大きく以下の 4 つの研究項目に分類される。

1）個人特徴顔モデリングに関する研究

個人特徴を有する顔合成のため、3 次元レンジスキャンした顔の立体形状データから、頭髪を含む頭部モデリングを実施する。図 7 に新旧のキャラクタを示す。図 7（左）は、フューチャーキャストのヘルメットのある頭部モデル、図 7（右）は、新しい頭髪を有する頭部モデルである。

髪型を実際の来場者から計測してモデル化する手法の開発が必要であるが、現時点でクオリティの高い頭髪モデリング手法は存在しない。一方、

頭髪は登場人物の配役特徴を表すものであり、また頭髪バリエーションも典型パターンに分類されることも多いことから、複数の候補の中から来場者の好みに合わせて選択する方式も考慮している。この場合、選ばれた頭髪モデルは顔モデルとの大きさの整合がとられ、頭皮モデルに対して最適な大きさが決定される。モデルは事前に作りこみできるため、あとはリアルタイムで運動制御できる頭髪の質感表現が必要である。現在は、複数の短冊モデルを頭部のレイヤー毎に頭皮に植え付け、この短冊に頭髪のテクスチャをα値にしたがってマッピングすることで実現している。

表情合成の個人特徴に関しては、現在表情筋モデリング手法を応用して個性の表現を試みており、表情筋バネモデルの配置位置および物理特性を個人ごとに最適化することで、できるだけ個人の特徴を有する表情合成方式について検討を行っている。また多波長 BRDF 計測を実施することにより個人の皮膚の質感の再現に関しても検討している。



ＦＣＳオリジナル　　頭髪を含むモデル

図７　　新旧キャラクタモデルの比較

２）個人特徴を有する全身モデリング

個人の身長や体格を反映したモデルとして、全身のキャラクタモデルを作成中である。図８にキャラクタモデルのおおよその体格の分類を示している。

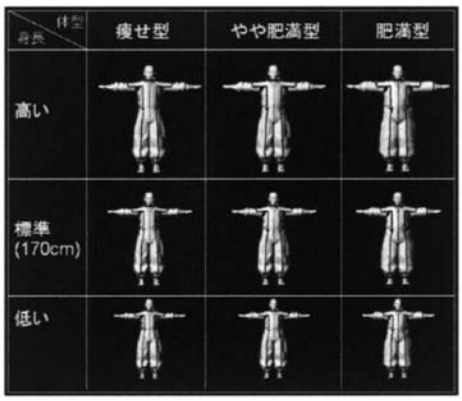


図８　体格の異なるキャラクタモデル

実際には、首の太さや頭部の大きさも異なるため、レンジスキャンした顔モデルの整合が課題である。また、動作についても個々の典型的な体格ごとに個性が反映さえるように制御されなくてはならない。

３）個人特徴を有する動作モデリング

歩容動作による個人認証の研究が行われている。これはシルエット画像に対して空間周波数特徴解析をすることにより個人の判別するものである。研究プロジェクトでは、まず大規模の個人歩容データベース構築を行っている。

これは被験者がトレッドミル上を様々な速度で歩行もしくは駆け足を行っている際に、対象を取り囲む複数の方向からカメラ撮影したものである。

一方、同時に個人の歩容動作をモーションキャプチャにより取得し、各関節角度の時間的な推移の特徴を解析した。図 10 にモーションキャプチャの様子（右）と、その際の肩と腰の回転角度の波形を示した（左）。この結果、動作の個人特徴は関節角度の基本周波数とその振幅および関節ごとの位相のずれによってモデル化できることが示唆された。実際には、詳細の動きデータベースの中からシルエットベースで近い動作を選択し、キャラクタに適用することで、個人特徴を表現する手法について検討している。



図９　大規模歩容データベース

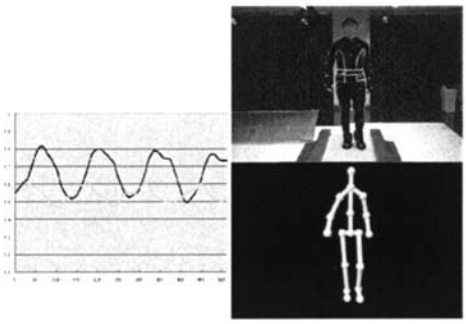


図 10　歩容モーションキャプチャ

４）個人特徴を有する声優音声生成技術

声の個人特徴表現に関しては、そもそも他人が聞いていて感覚的に個人の声に近いと感じる特徴が音声のどの物理パラメータとして表現できるの

かを明らかにする必要がある。音声合成のクオリティは劇場用には不向きなため、予め収録した多くのバリエーションの声優音声から、対象となる個人の特徴を反映するハイクオリティな声を生成しなくてはならない。

5. 役者中心仮想環境再現技術

「ダイブイントゥザムービー」のめざす２つ目の機能として、役者や映像カメラマンの立ち居地や収録環境をそのまま忠実に再現する３次元仮想没入空間の実現がある。映像を視聴者として第三者の視点で映像鑑賞するだけでなく、役者視点での映像鑑賞という新しい映像技術を提案し，登場人物やカメラマンの視点から、周りの状況を自由に鑑賞・体験できるシステムを実現する。

本研究テーマは、３つに分類できる。

１） 頭部装着型役者中心映像撮影システム

配役の立ち位置から、頭部を取り囲む３６０度全周囲の映像を目線を中心として、高画質で記録するヘッドマウント型カメラシステムを製作中である。これは４台のHD画質の軽量ハンディカムを曲面ミラーを介して全方位をくまなく撮影するものである、演算によって、およそ４８００画素×８００画素のパノラマ動画の取得および再生が実現できる予定である。

２） ３次元音場収録・再現技術の確立

役者もしくはカメラマンの置かれる環境において、本人の体験した音響空間を再現する技術は、没入環境を構築する上で非常に重要な要素となる。

図11に音場再現装置の概観を示す。体験できる人数は１度に１人であり、パーソナルな空間を実現する。複数のスピーカーアレーから構成される。またスピーカー配置と対応した３次元音場収録システムも実現しており、収録環境をそのまま再現することが可能である。今後は、１）のパノラマ映像との同時提示装置の実現が課題となる。

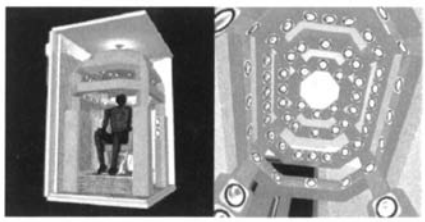


図11　音場再現装置の概観

6. おわりに

本研究プロジェクトの成果公開の機会として、本年度は、2008年2月10日（日）,11日（月）に、2006年度と同様に日本科学未来館にてアウトリーチ活動を予定している。今年度は、3つの大きな展示と要素技術の展示を計画している。

１つ目は、進化版のフューチャーキャスト体験である。、昨年のアウトリーチの際、声優の声が固定されていたが、今年度は、話者の発話音声特徴に応じて台詞音声の入れ替えを行う。

２つ目は、役者中心仮想環境再現に関する展示であり、頭部装着型全方位撮影システムを利用して、高画質のパノラマ映像をデモンストレーションする。また、試作した音場再現装置のデモを行い、実際の音響コンテンツを体験してもらう。

このほかに要素技術の展示を実施し、子供達に分かりやすく最先端技術を伝える予定である。

また今年度も体験した来場者を対象にアンケートを実施し、2008年度の最終年度への改良方針を決めて、研究を加速させる予定である。

研究プロジェクトの詳細は、下記のホームページを参照されたい.

http://www.diveintothemovie.net/jp/index.php

謝辞：本研究プロジェクトは、文部科学省 科学技術振興調整費「重要課題解決型研究等の推進」プログラムでの平成１８年度採択課題 「新映像技術ダイブイントゥザムービーの研究」により進められているものです。