

GAYAIT : 個と群衆が共存するインタラクティブシステム

水野 慎士[†] 平野 良介[†] 堤 幸彦[‡]

本論文では、同時に再生されている多数のビデオ素材の再生状態をユーザが対話的に変化させていくことができるインタラクティブメディアシステム"GAYAIT"を提案する。システムの操作はユーザの動作を Web カメラで取得して行う。システムのビデオ素材としてメッセージビデオを用いた場合、同時再生しているときには雑踏のワイワイガヤガヤという状態だが、ユーザ操作によって一つのビデオ素材を選択すると個人のメッセージが浮かび上がって聞こえてくる。歌唱シーンのビデオ素材を用いた場合には、全員の合唱合奏と各個人の独唱独奏を対話的に切り替えて楽しむことが可能である。GAYAIT は個と群衆の共存とその瞬時の切り替えを楽しむことができるシステムである。

GAYAIT : An Interactive Media System with Individual and Whole characters

SHINJI MIZUNO[†] RYOSUKE HIRANO[†] YUKIHIKO TSUTSUMI[‡]

In this paper, we propose "GAYAIT": an interactive media system in which the user can interactively control a large number of video clips played simultaneously. The user's motion is taken by a webcam and used to control the system. As using video clips in which a person is talking some messages, the system synthesizes an atmosphere like a crowd by playing each video clip simultaneously, and when the user picks up one of the video clips, the saying of the clip becomes clear and could be distinguished. As using video clips in which a person is singing a song, the user can enjoy a vocal solo of each people and their chorus interactively. The user can experience circumstances and a change of individual and whole characters in the system.

1. はじめに

近年のコンピュータやネットワークの技術の発達に伴い、画像処理技術、音声処理技術、ロボット技術などを用いたメディアアートが盛んになっている。特にユーザの動作に対してリアルタイムに反応するインタラクティブメディアは、広告媒体やコミュニケーションツールとしての活用が提案されており、ますます注目を浴びてきている^{1),2),3)}。

そこで本研究では、新しいインタラクティブメディアを実現する映像音声提示システム"GAYAIT"を提案する。GAYAIT は複数のビデオ素材を同時に取り扱いながら映像と音声をユーザに提示する。ビデオ素材を同時に扱うメディアアートとしては、ビデオアートの先駆者であるナム・ジュン・パイクが発表した、多くの TV モニターを並べて映像を同時に再生する作品⁴⁾や、液晶モニターで制作したクリスマスツリーを展示した"AQUOS Experience in the Grand Central 2008"⁵⁾などが知られている (図 1)。GAYAIT はこれらの作品

にインタラクティブ的な要素を加えたメディアアートとも言える。

GAYAIT では一つのディスプレイ内で数十本のビデオ素材が再生される。通常の状態では、全ての映像と音声は同じような大きさで同時に再生されており、全体で一つであったり雑踏であったりする。しかし、ユーザの動作に応じてディスプレイ中の一つのビデオ素材が選択されると、その映像や音声が強調されて含まれる個性が浮かび上がってくる。このように、GAYAIT は個と群衆の共存と瞬時の切り替えを対話的に体験できるシステムである。本稿では GAYAIT の実現方法や応用例を示す。

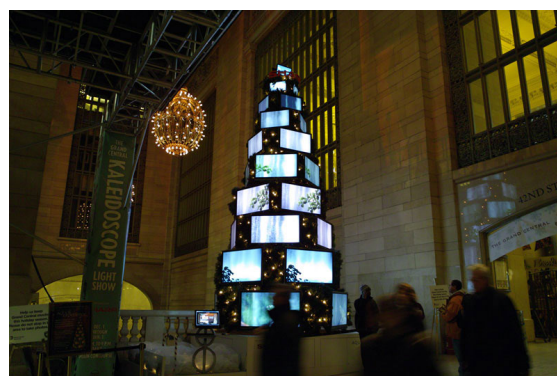


図1 AQUOS Experience in the Grand Central 2008.

[†] 愛知工業大学情報科学部
Faculty of Information Science, Aichi Institute of Technology
[‡] 株式会社オフィスクレッシェンド
Office Crescendo Inc.

2. GAYAIT の概要

GAYAIT の実現のために最低限必要な機材はパソコンと Web カメラである。またユーザの動きを追跡するためにカラーボールも使用する。パソコンは Web カメラからの映像の解析、ビデオ素材コントロール、提示用 CG 映像の生成を行い、生成された CG 映像はプロジェクタなどの大画面に投影される。ユーザはカラーボールを持って投影された映像の前に立ち、ユーザの動作は Web カメラで撮影される。図 2 に GAYAIT の展示状況を示す。

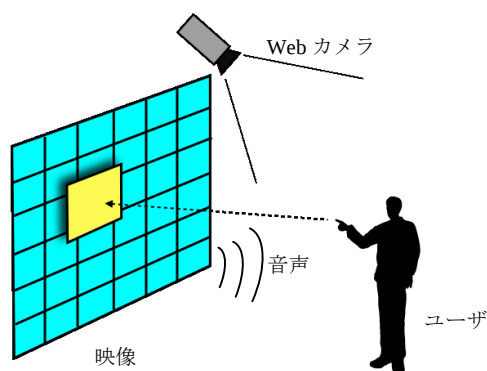


図2 GAYAIT の展示状況.

生成され投影された CG 映像中では、パソコン内に保存されたビデオ素材が CG 空間中に配置されて同時に再生されている。同時に扱うことができるビデオ素材数は 25～50 本程度で、その内容は例えば人がカメラの前でメッセージを話している映像である。そして初期状態ではすべてのビデオ素材の映像と音声は同じような大きさで再生されている。そのため、各ビデオ素材は大衆を構成する要素の一つに過ぎず、メッセージの内容を個別に判別することは難しく、全体としてはまるで雑踏の中にいるような「ワイワイガヤガヤ」という映像と音声となる。図 3 に GAYAIT の初期状態の画面を示す。



図3 GAYAIT の初期状態の画面.

Web カメラはユーザの動作を撮影しており、ユーザが持つカラーボールを検出するとカラーボールの追跡を開始する。そしてカラーボールの検出位置は CG 映像中のポインタ座標に変換される。ユーザにとってカラーボールはマウスのような役割である。ポインタ座標に最も近い位置に配置されているビデオ素材が選択されたことになる。

一つのビデオ素材が選択されると、選択された素材の CG 映像と音声は拡大される。それと同時に、選択されなかったビデオ素材の音声は低減される。その結果、選択されたビデオ素材が強調されて再生されるようになり、選択されたメッセージの内容は明瞭に認識することができ、大衆の中から個性が浮かび上がってくる。なお、ポインタ座標からの距離に応じて、各ビデオ素材の映像と音声の大きさは変化する。図 4 にビデオ素材が選択された様子を示す。

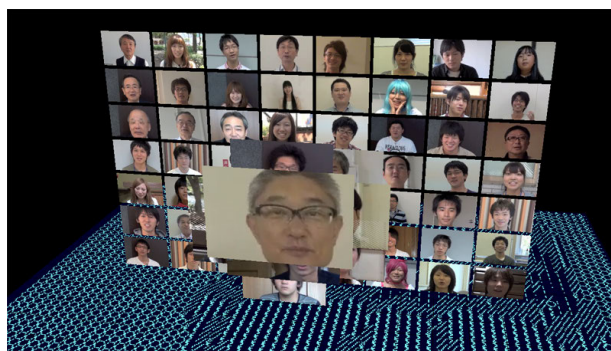


図4 ビデオ素材が選択された状態の画面.

ユーザの動作中は常にボールの追跡を行っており、逐次ビデオ素材の選択と CG 映像生成、音声再生を行う。そのため、ユーザの動作に応じてリアルタイムで対話的に映像と音声に変化する。

ビデオ素材は GAYAIT 動作中に逐次追加することが可能である。ビデオ素材数が最大同時再生数を超えた場合には、システムが定期的に取り扱うビデオ素材を変更していく。

3. GAYAIT の実現方法

GAYAIT の実現には、CG 映像・音声の生成、Web カメラの映像解析、Web カメラの解析結果に基づく映像・音声のコントロールなどの処理が必要となり、C++を用いて各処理が実装されている。以下のに各処理の詳細を述べる。

3.1 空間の構築と CG 映像と音声の再生

GAYAIT の映像は三次元 CG に基づいて生成されている。CG 空間には多数の長方形ポリゴン配置されて

いる。これらは仮想ディスプレイの役割を持ち、それぞれの中でビデオ素材が個別にループ再生される。そしてユーザ動作に応じて長方形ポリゴンの座標やサイズが対話的に変化する。

ビデオ素材は映像データと音声データに分けられてファイルとして保存されている。各長方形ポリゴンには個別のビデオ素材が関連付けされており、ビデオ素材ビデオ素材が長方形ポリゴンより多い場合には定期的に関連付けを変更する。

各ビデオ素材の映像のフレームは画像として抽出される。そして抽出されたフレーム画像は CG 空間中の対応する長方形ポリゴンにテクスチャマッピングの技法を用いて貼り付けられる。すべてのビデオ素材に対してこの処理をフレームごとに行うことで、長方形ポリゴンへのビデオ映像の貼り付けを実現している。

ビデオ素材の音声は音量に関するパラメータを持たせている。初期状態ではすべてのビデオ素材の音量レベルはほぼ同一であり、ユーザ動作に応じて各ビデオ素材の音量パラメータを変化させて特定の映像素材の音声を調整する。

3.2 Web カメラの映像解析

現状の GAYAIT では、ユーザが緑色や青色のカラーボールを持って動作することでシステムを操作する。Web カメラでユーザの動作を撮影して、ユーザの持つカラーボールを抽出することで、システムを操作するポイント座標を決定する。

カラーボールの抽出手順は以下の通りである。

<色に基づく領域の抽出>

Web カメラから取得したユーザ動作の映像を RGB 形式から HSV 形式に変換して、色相および明度に基づいて特定の色成分を持つ領域を抽出する。抽出例を図 5 に示す。



図5 色成分による領域抽出。

<小領域の除去>

色情報に基づく抽出結果では非常に多くの小領域が

ノイズとして含まれている。そこで領域収縮膨張処理を施すことで、小領域を除去する。処理例を図 6 に示す。



図6 領域収縮膨張処理による小領域の削除。

<面積に基づくボール領域の選択>

ここまでの処理で残ったすべての領域について面積の計算を行い、最大面積を持つ領域をカラーボール領域と認定する。なお、得られた領域の面積があるしきい値以下の場合にはボール領域は存在しないと判断する。処理例を図 7 に示す。

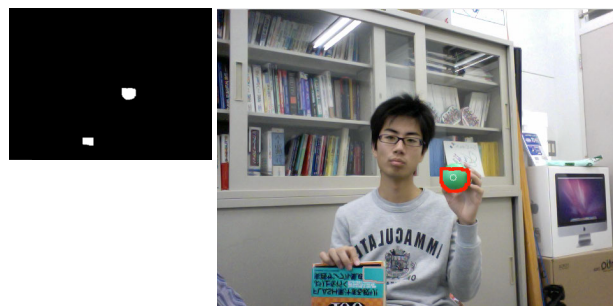


図7 カラーボール領域の推定。

<領域輪郭に基づく領域中心点の計算>

カラーボール領域の輪郭追跡を行って、 n 個の輪郭線上の点を抽出して、輪郭開始点 (0 番目) および輪郭中間点 ($n/2$ 番目) の中点をカラーボール領域の中心点とする。本手法で得られる座標は実際の中心点とは一致しないが、GAYAIT の操作には大きな問題にはならない。中心点の決定手順を図 8 に示す。

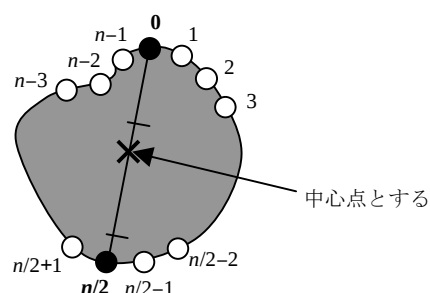


図8 カラーボール領域の中心座標の決定方法。

得られたカラーボール領域の中心点に基づいてシステム操作ポイントの二次元座標が決定される。

3.3 映像・音声のコントロール

前節で述べたように、GAYAIT ではビデオ素材を再生する長方形ポリゴンが三次元空間に配置されている。そしてユーザ操作によって決定したシステム操作ポイントの二次元座標が GAYAIT の三次元空間に投影される。この投影点が長方形ポリゴン上にあるとき、この長方形ポリゴンが選択されたことになる。

一つの長方形ポリゴンが選択されると、そのポリゴンのサイズが拡大されて、三次元座標も視点方向に移動する。その結果として、その長方形ポリゴンにマッピングされたビデオ素材の映像が拡大・接近されて表示される。また選択された長方形ポリゴンの周囲にある長方形ポリゴンも、投影点からの距離に応じてサイズと座標が変更される。

長方形ポリゴンが選択されたときは、その長方形ポリゴンに対応したビデオ素材の音声再生の音量も増大させる。そして、それ以外のビデオ素材の音量は投影点からの距離に応じて減少させる。

以上の処理を逐次行うことで、通常時は同じようなサイズと音量で再生されているビデオ素材が、ユーザ操作によって一つのビデオ素材を選択すると、その映像と音声が拡大したり、その他のビデオ素材の再生をコントロールすることにより、選択したビデオ素材が強調されて再生されるようになる。図 9 にビデオ素材の選択の様子を示す。ポインタ（×印の位置）で一つのビデオ素材が選択され、その周囲のビデオ素材もある程度の影響を受けている。

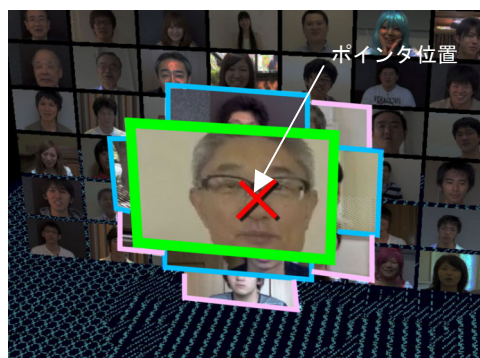


図9 ポインタの位置によるビデオ素材の選択と周囲への影響。

4. 実験

提案した GAYAIT システムを実装して、2 種類のビデオ素材を用いて実験を行った。使用したパソコンは MacBook Pro (OS 10.6.8, Core i7 2.8GHz, 8GB メモリ) で、システムは C++を用いて開発した。ビデオ

素材データは映像サイズが 160×120 画素で、音声はステレオ 48KHz である。また、画面表示のフレームレートは 10 フレーム/秒となっている。現状のシステムで同時に再生できるビデオ素材は最大 56 本である。ユーザは緑色のボールを持ってシステムの操作を行う。ユーザ動作の取得には 300 万画素の Web カメラを用いた。図 10 にシステムに使用した主な機材を示す。

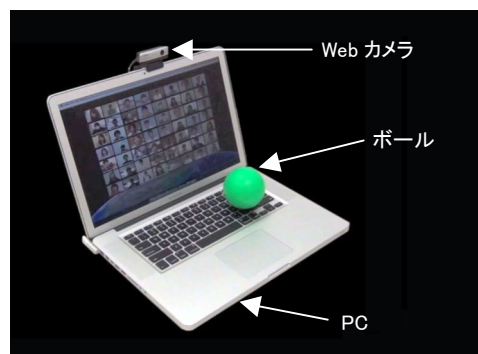


図10 GAYAIT で使用する主な機材一式。

4.1 メッセージビデオ素材を用いた実験

最初にビデオ素材として個人がメッセージを話しているシーンを収録して用いた。ビデオ素材の総数は 85 本で、その中から 56 本をランダムで選択して使用する。メッセージの長さは 15 秒から 7 分と様々であり、話している内容も多種多様である。GAYAIT 上では各ビデオ素材を独立してループ再生するため、メッセージの進行状況はバラバラとなる。また一定時間ごとにビデオ素材の入れ替えを行う。ビデオ素材入れ替えの様子を図 11 に示す。

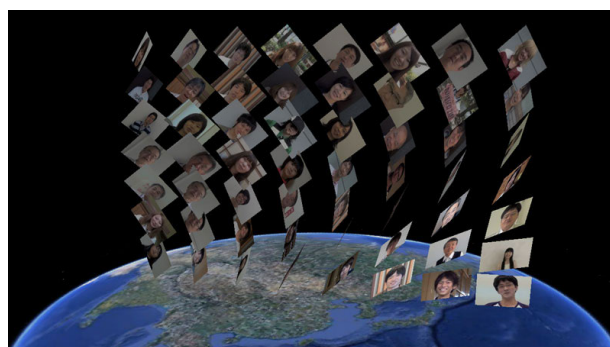


図11 ビデオ素材入れ替えの様子。

メッセージビデオ素材が選択されていない状態では、異なる内容のメッセージが同時に同じような音量で再生されるため、まさに雑踏の中にいるような状況となった。所々で内容が聞き取れるメッセージもあるが、ほとんど何を話しているか判別不明であり、没個性の

状態となっている。

そして、ユーザが Web カメラの前にボールをかざすと即座にボールが認識される。そしてボールの位置に応じて一つのメッセージビデオ素材が選択されると、瞬時にその映像サイズと音量が拡大されるとともに、その他のビデオ素材の音量が低下して、選択された素材のメッセージが強調された。図 12 に選択されるメッセージビデオ素材が徐々に移動していく様子を示す。



図12 選択するメッセージビデオ映像が移動する様子。

GAYAIT でメッセージビデオ素材を用いた場合、ユーザの操作に応じて次々にビデオ素材が選択されていくと、群衆の雑踏に対して虫眼鏡を使ったように映像と音声が拡大されて、メッセージを通じた個性が浮き上がってくるような感覚となった。

メッセージビデオ素材を用いた GAYAIT を大型プロジェクトに投影して一般の人々に公開して実験を行った。実験では 10 代から 60 代まで 50 人以上が GAYAIT を体験して、23 人からはアンケート回答も得られた。ほとんどの人が生成される映像音声に興味深く、非常に面白いという感想であった。図 13 に大型プロジェクトを用いた実験の様子を示す。

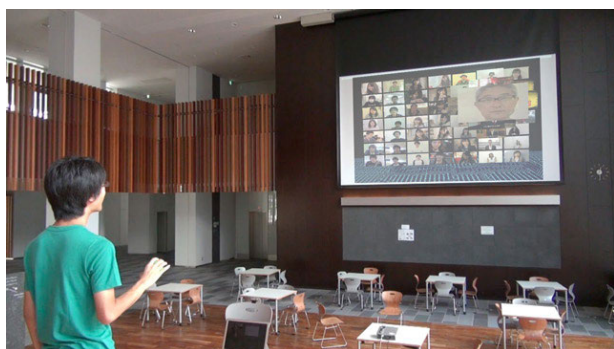


図13 大型プロジェクトを用いた実験。

4.2 歌唱ビデオ素材を用いた実験

次にビデオ素材として個人が歌ったり演奏しているシーンを用いた。ビデオ素材は 25 本で、23 本は歌を

歌い、2 本は楽器（クラリネット、リコーダー）を演奏している。全員が同じ歌（Happy Birthday to You）の同じ伴奏に沿っているが、収録は別々に行っている。ビデオ素材の長さは全て 30 秒で、すべて同時に再生を開始して、2 回目以降はループ再生を行う。

歌唱ビデオ素材が選択されていない状態では、全員の歌が同期されて同時に同音量で再生される。ビデオ素材には男性、女性、楽器が混じっており、歌唱力もばらつきがあるため、必ずしも洗練されているとは言えないが、まるで合唱合奏のような音声が再生された。

そしてユーザのボール操作によって一つの歌唱ビデオ素材が選択されると、瞬時に合唱から独唱独奏に切り替わるような状態となる。操作を続けて選択するビデオ素材を次々に切り替えていくと、一つの歌を別々の人が次々と出現しながら歌ったり演奏したりするという、他にはあまり例がないような映像が対話的に生成された。

GAYAIT で歌唱ビデオ素材を用いた場合、合唱合奏での集合の個性と、独唱独奏での個の個性を対話的に切り替えて感じることができた。図 14 に GAYAIT で歌唱ビデオ素材を使用した例を示す。



図14 歌唱ビデオ素材を用いた GAYAIT。

5. 考察

GAYAIT の特徴である個と群衆、個性と無個性の共存と対話的な切り替えは様々な場面での活用が考えられる。例えば、結婚式などのお祝いメッセージを GAYAIT で用いた場合には、全員が新郎新婦へお祝いの言葉を話しているが、同時に再生しているために個々のメッセージ内容はわからない。ただし、全員がにこやかな表情をしており、幸せを感じることができる。そして一人の映像が選択されると、その人のメッセージが浮かび上がり、より一層個人のメッセージが

印象深く感じられると思われる。同様の効果は広告媒体でも期待できる。

また、美術館や博物館での収蔵品の新たな展示方法、合唱や合奏の新たな視聴方法、教育分野での活用も考えられる。例えば、地域振興のために制作されたカルタを GAYAIT に適用することを計画している⁵⁾。

6. まとめ

本稿では、個と群衆の共存とユーザ操作に応じた瞬時の切り替えを実現するインタラクティブ映像音声提示システム"GAYAIT"を提案した。システムの概要や特徴、実現方法を述べると共に、システムを実装して実験を行った。実験の結果から、GAYAIT は様々な場面で活用できる可能性があることがわかった。

今回はビデオ素材としてメッセージと歌唱のシーンと音声を用いたが、今後は他にどのような素材が適用できるか実験を行っていく。また Kinect などによるユーザ動作取得手法を適用して、より多くの場面で活用できるようにシステムを改良するつもりである。

謝辞 本研究の遂行にあたって実験協力や熱心な討論を頂いた愛知工業大学情報科学部水野慎研究室および澤野研究室の皆様に感謝する。またビデオ素材の撮影に協力して頂いた多くの皆様に感謝する。本研究の一部は科学研究費補助金（23500139）による。

参 考 文 献

- 1) 大島直樹, 岡澤航平, 本田裕昭, 岡田美智男: TableTalkPlus: 参加者の共同性や社会的なつながりを引き出すアーティファクトとその効果, ヒューマンインタフェース学会論文誌, Vol. 11, No. 1, pp. 105-114 (2009).
- 2) 椎尾一郎, 山本吉伸: コミュニケーションツールのための簡易型 AR システム, コンピュータソフトウェア, Vol. 19, No. 4, pp. 2-9(2002).
- 3) Google Chrome Music Mixer: <http://www.morewithgoogle.jp/musicmixer/>.
- 4) 李容旭: 電子メディア時代の映像表現と創造性 (1)ーナムジュン・パイクの場合, 東京工芸大学芸術学部紀要 13, 13-18 (2007).
- 5) まちの会所 hanare: <http://kin2hanare.jp/>.