

招待論文

動画視聴をベースとした非同期コミュニケーションシステム WakWak Tubeの開発

坂内 祐一^{1,a)}

受付日 2018年1月23日

概要: YouTubeなどのオンデマンド動画サービス視聴者の非同期コミュニケーションを実現するシステム WakWak Tube を開発した。このシステムは、YouTube の動画を視聴しているユーザの動きを Kinect により取得して、アバタとしてリアルタイムに表示すると同時に記録する。YouTube 動画再生が繰り返されるたびに、視聴中のユーザのアバタと過去のアバタが同期再生される。またユーザアバタの選択表示機能や他のアバタを移動させるというインタラクション機能を実装し、他ユーザとの一体感を醸成してユーザの臨場感の増進を図った。さらにインタラクションの体験を共有する機能を設けることで、非同期コミュニケーションを促進することが期待できる。

キーワード: 非同期コミュニケーションシステム, 動画視聴, 身体情報, インタラクション

WakWak Tube: An Asynchronous Communication System Based on Video Viewings

YUICHI BANNAI^{1,a)}

Received: January 23, 2018

Abstract: We developed an asynchronous YouTube viewing system called WakWak Tube, which captures the movement of the viewer's body using a Microsoft Kinect sensor as a YouTube video plays, and displays the viewer as an avatar in an audience area of the screen in real time. A new viewer can look at other avatars as well as his or her own avatar synchronously with the video playback. In this paper, we propose a system that enables viewers to foster a sense of connection with other avatars by selecting avatars recorded in the past and by interacting with the other avatars. We aim to expand this system as an asynchronous communication system based on video viewing.

Keywords: asynchronous communication, video viewings, kinetic information, interaction

1. はじめに

テレビ (TV: Television) が登場した 1950 年代, 高価な TV 受信端末は茶の間に置かれ, 家族全員で感想を述べ合ったり番組の情報を共有したりしながら視聴するのが一般的だった。TV 端末の低価格化にともない, TV は個人の所有となり視聴形態も個別化していく一方で, コンピュータやインターネット技術の進展により, ネットワーク上のサー

バに蓄積されている動画を任意の時刻に再生するオンデマンド動画配信サービスが一般的になった。YouTube [1] はその代表例である。TV 放送においては, 一方向に電波を発信するだけの従来型放送から, 視聴者からのレスポンスを番組内容に反映させるなどのインタラクティブな機能や, 番組の視聴者間での情報共有やメッセージの交換などのコミュニケーション機能の検討が行われてきた。後者に重点を置き TV とコンピュータの融合を図るシステムは, 海外ではソーシャル TV という範疇に分類され, 2000 年代に入ってからさまざまな研究やサービスが行われている [2], [3]。これらのシステムでは, 内容に関するコメントをテキストや音声のチャットで交換する機能や, ビデオク

¹ 神奈川工科大学
Kanagawa Institute of Technology, Atsugi, Kanagawa 243-0292 Japan

^{a)} bannai@ic.kanagawa-it.ac.jp

リップをシェアしたり番組を推薦したりする視聴者間コミュニケーション機能が、TV 視聴に同期して利用することを前提としているものが多い。国内に目を向けると、上記のような視聴者間コミュニケーションの大部分が SNS によって行われている [4]。またオンデマンドの動画配信サービスとして人気を集めているニコニコ動画 [5] では、視聴者が非同期に動画に対して投稿したコメントが動画再生時間に同期して画面上に表示され、コメントによる盛り上がり共有しているという特徴がある [6]。

本研究では、ニコニコ動画のような動画配信サービスで提供されている非同期の視聴・コメント付与・コメント再生などのコミュニケーションを、より一般化させるためノンバーバル情報を導入する。WakWak Tube は、YouTube の動画を視聴しているユーザの動きを Kinect により取得して、アバタとしてリアルタイムに表示すると同時に記録するシステムである。YouTube 動画再生が繰り返されるたびに、視聴中のユーザのアバタと過去のアバタが同期再生されるので、ユーザの臨場感や他ユーザとの一体感が増進されることが期待できる。また WakWak Tube では、このような非同期のアバタから生成される共有空間において、ユーザアバタが他のアバタを移動させるというインタラクション機能も実現した。本論文では WakWak Tube のこれらの機能を紹介するとともに、非同期アバタ間で可能なインタラクションを取り上げ、これらのインタラクションがさまざまな体験をもたらすこと、さらにこの体験を共有する非同期コミュニケーションの可能性について述べる。

2. 関連研究

TV 視聴をベースとして、ユーザ間でのコミュニケーションをサポートする機能を実現した初期のシステムに AmigoTV [7] がある。このシステムでは、TV 映像を視聴しているユーザはアバタ表示され、音声や感情を表すエモティコンを含むテキストメッセージを伝送することで、ユーザの参加意識を高めることを目指している。

Harrison らの CollaboraTV [8] は、非同期に遠隔地の人とともに映像を共有しているかのような体験を与えるシステムである。現在のユーザが見ているのと同じ映像を過去に視聴したことのあるユーザがアバタとして映像鑑賞画面の下部に表示され、ユーザのポジティブ・ネガティブの感情をアバタの動作として反映させたり、感情の統計を興味プロフィールとして表示したり、ユーザのコメントをアバタからの吹き出しで表示したりする機能で、ユーザとの一体感を醸成し動画をより楽しく感じさせることを狙っている。代蔵らの開発した ExciTube [9] では、異なる場所にいるユーザ同士が興奮情報を共有することで、実際に他者が同じ場所にいるかのような体験を提供している。ユーザの興奮度は手掌に電極を装着し SCR (Skin Conductance Response: 皮膚コンダクタンス反応) で計測する。ニコニコ

動画と比較して必ずしもユーザ反応情報が増えているとは言いがたいが、SCR という感情が直接現れる生体信号を用いているところに特徴がある。

吉田ら [10] は、動画の前に立つ観客の身体動作を“ノリ”と定義し、動画上に身体操作を重畳表示するシステムを開発した。この“ノリ”情報を動画の時間軸上に載せ、複数ユーザ間で非同期に共有し、異なる時間・空間にいる他のユーザとの一体感を得ることを狙っている。マイクロソフトのキネクト (Kinect) を用いて観客の骨格情報を取得し、動画上に棒状のアバタとして表示することで“ノリ”を可視化し、臨場感から受けることができる感情である“一体感”が得られたかを検証している。学会会場で展示した際のアンケートから、画面内で自分自身の骨格が認識困難であったこと、過去や現在のノリとのインタラクションができていなかったことで、一体感を感じられなかったと結論付けている。

川井らのシステム [11] では、ニコニコ動画のような動画共有サービスに、視聴者間でより一体感を持たせることを狙い、動画に重畳されたコメントをクリックすることでコメントの評価を行うインタフェースを提供している。この評価に応じてコメント表示にエフェクトを付け、エフェクトの視覚効果によりコメントに対する評価が容易に確認できるようにしている。

Audience Silhouettes for TV [12] は、視聴者の体の動きをリアルタイムで可視化して、TV 映像に重畳することができる同期型システムである。Kinect の深度センサを用いて視聴者の動きと姿勢をキャプチャして、シルエットとして表示している。このシステムは視聴者のシルエットを共有することで、ノンバーバル情報を伝え視聴者間で一体感を醸成することを狙っている。しかしながらシルエットを表示できるユーザ数や表示領域に限りがあるほか、ユーザ間のインタラクションについて議論がなされていない。

3. WakWak Tube

3.1 WakWak Tube の概要

WakWak Tube の利用法を図 1 に示す。キネクト v2 (Kinect v2) が設置されている PC で、ユーザが視聴する

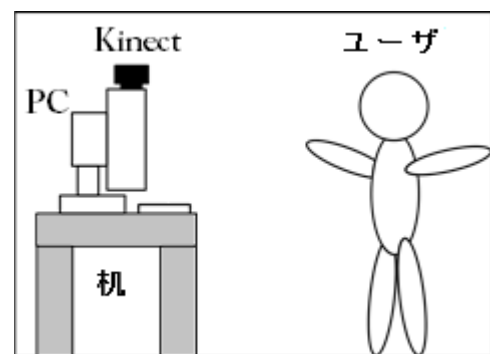


図 1 WakWak Tube の構成

Fig. 1 Configuration of WakWak Tube.

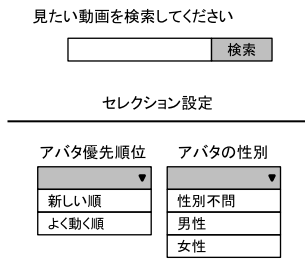


図 2 検索条件設定画面

Fig. 2 Panel for search condition setting.

YouTube の動画を選択すると、動画が再生されている間 Kinect v2 の深度センサによりユーザの骨格データが取得される。ユーザが PC モニタの前に立ち、YouTube 動画を視聴しながら体を動かすと、ユーザアバタのアニメーションが、YouTube 動画再生領域の下部にリアルタイムで表示される。このときユーザアバタの骨格情報はシステムに記録されていくので、YouTube 動画再生時には、視聴中のユーザアバタと蓄積された過去アバタアニメーションが同時に再生される。

3.2 WakWak Tube のインタフェース

WakWak Tube 使用にあたって、ユーザはログインパネルで、ID を入力し性別を選択してシステムにログインする。すると図 2 に示すような検索条件設定画面が表示される。テキスト入力欄に YouTube 動画検索のためのキーワードを入力して検索ボタンを押すと、キーワードに合致した動画がリストアップされる。画面下のセレクション設定では、自身のアバタと同時に表示する過去アバタを蓄積されたものから選ぶ基準を指定する。アバタ優先順位メニューの「新しい順」を選択すると、蓄積時刻が最近のアバタから順に選ばれ、「よく動く順」を選択すると活発に動いたアバタの順に表示される。アバタの性別メニューでは、表示させるアバタが「性別不問」・「男性」・「女性」を選択することができる。以上の設定を終え検索ボタンを押すと、YouTube 動画タイトルと要約のリストが検索結果として表示される。

この画面で所望の動画を選択すると図 3 に示すような WakWak Tube 画面に進み、動画再生が開始されると同時に Kinect v2 がユーザ骨格情報のトラッキングを開始し、動画再生終了までユーザの動きはアバタアニメーションとして動画と同期して表示される。

WakWak Tube 画面の上部にはステージに見立てた YouTube 動画表示領域が、画面下部には観客席をイメージした視聴者領域があり、アバタはすべてこの視聴者領域に現れる。視聴者領域の中央の赤いアバタは、現在視聴中のユーザアバタであり、異なる色のアバタは過去に蓄積されたものである。視聴者領域の大きさが限られているので、



図 3 WakWak Tube 画面

Fig. 3 WakWak Tube screen.

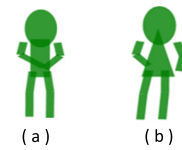


図 4 男性アバタ (a) と女性アバタ (b)

Fig. 4 Male avatar (a) and female avatar (b).



図 5 押しのけ機能

Fig. 5 Screenshot of the current user's avatar pushing away other avatars.

表示される過去アバタの数は 8 人までに制限されている。アバタを表示する際には衝突判定を行って、アバタが重ならないようにしている。アバタの性別が見分けられるようにアバタの胴体に異なる形状を持たせている。図 4(a) は男性アバタ、(b) は女性アバタである。

もし視聴中のユーザアバタが動くために十分なスペースがないときには、図 5 に示すように手で過去アバタを押しつけて隅へ追いやることができる (押しのけ機能)。また表示されている過去アバタから特定のアバタを選んで、自身のアバタの隣に配置することもできる (引き寄せ機能)。押しのけ機能の場合には、ユーザは手を開いたまま過去アバタに接触させ腕を伸ばすことで押しのけ動作が実現され、引き寄せ機能の場合には、ユーザは所望のアバタの位置まで移動して手を結んだ状態でアバタに接触すると、接触されたアバタはユーザに手を引かれた状態となり、ユーザの移動でアバタを一緒に移動することができる。

4. システムの実装

ユーザの骨格情報を取得するために、Windows SDK 2.0 for Kinect v2 を用いた。Kinect v2 は 6 人のユーザの 25 点の骨格情報をトラッキングすることが可能である。Microsoft .Net Framework に含まれる GUI 開発環境であ

る WPF をベースに、C#を用いて WakWak Tube のプログラムを開発した。

4.1 動画検索

動画検索はユーザがテキストボックスに入力したキーワードをもとに行う。キーワードの文字列をパラメータとして YouTube Data API の検索命令を実行すると、検索結果のリストが返される。このリストには、動画 ID、サムネイルの URL (Uniform Resource Locator)、動画キャプションなどの情報が含まれおり、サムネイルとキャプションをセットにして検索結果画面で一覧表示している。サムネイルとキャプションには動画 ID タグを付けてある。サムネイルまたはキャプションがクリックされ、そのイベントが検知されると図 3 の動画再生画面に遷移するが、その際紐付けされている動画 ID タグを再生画面クラスの引数に渡すことで動画プレーヤの動的生成を実現している。動画プレーヤは WPF の Web Browser コントロールで YouTube 組み込みプレーヤの URL に接続して表示している。

4.2 アバタ情報

アバタは Kinect の骨格追跡機能を利用して再生している。Kinect v2 では人体の 25 カ所の骨格座標を取得できるが、図 6 に示すように本システムではその中から 11 点を除いた 14 点を使用する。再生に用いる骨格情報の数を減らすことで処理速度を安定させるためである。骨格座標は 10fps で取得し、画面出力用の 2D 座標に変換してから配列に格納する。骨格座標配列は 2 次元で、1 次元目が動画フレーム番号、2 次元目が骨格の種類となっている。この配列要素をもとに頭の座標に楕円を、胴体中心の座標に長方形 (男性アバタ) または三角形 (女性アバタ) を描画し、四肢には関節座標を始点と終点に設定した線分を描画することで人型アバタとして表現している。

4.3 衝突判定

それぞれのアバタの表示領域を確保するために、アバタ領域を長方形として衝突判定を行っている。2 つの長方形領域の重なった部分を返す Rect クラスの Intersect 関数を用いて、重なった部分を確認し、重複領域の大きさの分

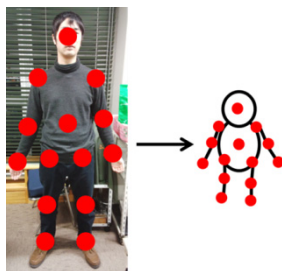


図 6 骨格点

Fig. 6 Joints used in WakWak Tube.

アバタ座標のオフセット値を加減算することで衝突を回避し、アバタの領域確保を実現している。衝突判定に使用する四角形は、押しのけ機能の場合はユーザアバタの手に、接近による回避の場合はユーザアバタの胴体に設置する。過去アバタにはどちらの場合であっても胴体に設置する。領域確保によってアバタ表示位置がずれるのは一時的なものであるため、ファイルには記録していない。

4.4 アバタ情報の記録と読み込み

本システムはユーザの身体動作を非同期で共有するため、トラッキングされた時系列の骨格座標、すなわちアバタの振舞いをファイルに保存している。アバタ情報の記録は WakWak Tube 画面での動画プレーヤ読み込み直後から動画の再生終了までの間行う。終了のタイミングは、経過時間と YouTube Data API により取得した動画の長さとを比較して計測している。

アバタ情報の記録には csv 形式を用いた。ファイル名は動画 ID.csv とし、動画 ID ごとにファイル名を分けて保存した。csv ファイルのヘッダには、動画を最後まで再生した場合に記録できる最大フレーム数を記述した。このフレーム数がアバタの骨格座標配列の 1 次元目の長さとなる。ヘッダの次の行からはアバタ情報が格納されている。アバタ情報は初めの行にユーザ ID と座標のオフセット値が入っており、この値はアバタの画面上での表示位置をずらす際に用いる。オフセット値があることで骨格座標を書き換えずにアバタの表示位置を動かすことが可能となる。その次の行から 14 種の骨格 2D 座標を記録している。2D 座標の情報はフレームごとに改行しているため、行数を見ればフレーム番号が分かるようになっている。ユーザ ID、座標のオフセット値、骨格座標リストをまとめたものが 1 人分のアバタ情報であり、次の行からは別のアバタ情報が入る。アバタ情報は動画を再生するごとに 1 人分ずつ増えていくため、ファイルの下層ほど新しい情報となる。記録はアバタのアニメーション間隔と同じ 1 秒間に 10 回の頻度で行っている。実際に作成した csv ファイルを図 7 に示す。見やすさを考慮し、行の一部を省略してある。0X, 0Y, 1X, 1Y というのは四肢の描画で使用する線分の始点と終点の X, Y 座標である。

4.5 データ構造

WakWak Tube で管理されているデータを木構造で示したのが図 8 である。システム情報はユーザ情報と再生された YouTube の動画 ID 情報からなる。ユーザ情報にはユーザ ID とパスワードが含まれ、YouTube 動画 ID 情報に保存されているアバタの数とアバタデータが含まれる。アバタデータは静的なデータと動的なデータに分けられる。静的データは、ユーザ名や性別などアバタを表示するときに使われるデータであり、動的データは、音声データ、フレー

	A	B	C	D	E
1	fnum3230				
2	id1	oX200	oY150		
3	0X64	0Y38	1X75	1Y60	2X86
4	0X64	0Y39	1X75	1Y60	2X86
5	0X64	0Y39	1X74	1Y60	2X85
6	0X64	0Y39	1X74	1Y61	2X84
7	0X64	0Y39	1X73	1Y61	2X85
2928	id2	oX350	oY150		
2929	0X64	0Y38	1X75	1Y60	2X86
2930	0X64	0Y39	1X75	1Y60	2X86
2931	0X64	0Y39	1X74	1Y60	2X85
2932	0X64	0Y39	1X74	1Y61	2X84
2933	0X64	0Y39	1X73	1Y61	2X85

図 7 骨格情報格納形式
Fig. 7 Joints data format.

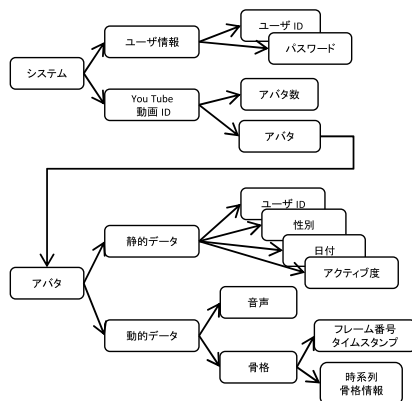


図 8 WakWak Tube データ格納形式
Fig. 8 Data structure of WakWak Tube.

ム番号、タイムスタンプ、および 4.4 節の骨格情報の時系列データであり、アバタアニメーションをリアルタイムに再生するために用いられる。

4.6 静的データ

4.4 節で述べたように、時系列の骨格データは処理速度を上げるため csv フォーマットで記述しているが、静的データは検索での利用およびシステムの拡張性を考慮して XML 形式を用いた。静的データはセレクション履歴のほか、アバタとユーザを紐付ける役割がある。図 9 の例では、ルート要素である StaticDataList は子要素として StaticData を持ち、アバタ id である avatarId を属性として持ち、子要素に SelectionSetting と AvatarProperty を持っている。SelectionSetting にはアバタの優先順位である Priority 要素と絞り込み処理用の性別を指定する Gender 要素がある。AvatarProperty 要素はユーザ id 属性を持っており、子に記録日時 DateTime 要素とアクティブ度を保持する ActivePoint 要素がある。アクティブ度は身体動作の記録が終了するまで算出できないため、ファイルを出力するのは動画終了時である。

4.7 LINQ to XML

静的データを検索するために、メモリ内 XML プログ

```
<? xml version="1.0" encoding="UTF-8" ?>
< StaticDataList >
  < StaticData avatarID="1" >
    < SelectionSetting >
      < Priority > newer </ Priority >
      < Gender > none </ Gender >
    </ SelectionSetting >
    < AvatarProperty userID = "0" >
      < DateTime > 2017/01/06 18:41:01 </
DateTime >
      < ActiveValue > 108.74 </ ActiveValue >
    </ AvatarProperty >
  </ StaticData >
  < StaticData avatarID = "2" >
    <!-- ..... -->
  </ StaticData >
</ StaticDataList >
```

図 9 XML 形式の静的データの例
Fig. 9 Example of static data.

ラムインタフェースである LINQ (Language Integrated Query) を用いた。データベースのクエリを C# のコードで記述して、XML やプログラムコード内の配列、リストに対して SQL のような問合せ処理を実行できる。

4.8 アクティブ度の算出

WakWak Tube のユーザは、YouTube の動画と同期再生されるアバタを、よく動く順に 8 体まで選択することができる。ここでは時系列骨格情報をもとに、動きの激しさを示すアクティブ度の算出について述べる。

フレーム i の胴体中心である SpineBase の座標を (x_{is}, y_{is}) 、骨格点 j の座標を (x_{ij}, y_{ij}) としたとき、画像原点と SpineBase との距離 $dist_os(i)$ と、SpineBase と各骨格点との距離 $dist_sj(i)$ は、式 (1)、(2) のようになる。

$$dist_os(i) = \sqrt{x_{is}^2 + y_{is}^2} \quad (1)$$

$$dist_sj(i, j) = \sqrt{(x_{ij} - x_{is})^2 + (y_{ij} - y_{is})^2} \quad (2)$$

方式 A では、体の中心である SpineBase 座標の標準偏差 $stdev(dist_os(i))$ と、SpineBase と骨格点の距離の変動を表す $stdev(dist_sj(i, j))$ を、骨格点ごとに計算する。これらの和をアクティビティ値 A とする。すなわち体全体の動きの激しさである SpineBase 座標の標準偏差と体の各部位の動きの激しさを表す SpineBase と骨格点の距離の標準偏差をアクティビティ値 A とするものである。A は式 (3) で求められる。

$$A = stdev(dist_os(i)) + \sum_{j=1}^{14} stdev(dist_sj(i, j)) \quad (3)$$

(for $1 \leq i \leq maxframe$)

方式 B では、各骨格点と画像原点との距離 $dist_{oj}(i, j)$ を式 (4) で求め、

$$dist_{oj}(i, j) = \sqrt{x_{ij}^2 + y_{ij}^2} \quad (4)$$

骨格点ごとの標準偏差の和をアクティビティ値 B とする (式 (5)). この値は体の構造を考えずに、各骨格点の動きを画像原点からの座標変化の標準偏差で表し、単純合計したものである.

$$B = \sum_{j=1}^{14} stdev(dist_{oj}(i, j)) \quad (5)$$

方式 C では、各骨格点座標のフレーム間差分を計算する. 骨格点 j の $(i-1)$ 番目のフレームと i 番目のフレームとの差分 $dif_{ij}(i, j)$ は、式 (6) で求められる.

$$dif_{ij}(i, j) = \sqrt{(x_{ij} - x_{i-1j})^2 + (y_{ij} - y_{i-1j})^2} \quad (6)$$

i 番目のフレーム間差分の合計は式 (7) で求められるので、合計値の平均を式 (8) で求めて、アクティビティ値 C とする.

$$dif_{i(i)} = \sum_{j=1}^{14} \sqrt{(x_{ij} - x_{i-1j})^2 + (y_{ij} - y_{i-1j})^2} \quad (7)$$

$$C = average(dif_{i(i)}) \quad (for\ 2 \leq i \leq maxframe) \quad (8)$$

方式 A, B, C は以下のようにまとめることができる.

- (A) SpineBase の標準偏差と SpineBase と各骨格点との距離の標準偏差の総和
- (B) 各骨格点座標の標準偏差の和
- (C) 各骨格点座標のフレーム間差分の和

5. システムの評価

5.1 アクティブ度の評価

3つのアルゴリズムで計算されたアクティブ度が主観評価とどう一致するのかを、図 10 に示す YouTube 動画に対して保存されたアバタを用いて実験した. 図 10 の (1) は 3 人の女性インストラクタがラジオ体操を行っている動画 (6 人の男性アバタと 3 人の女性アバタが保存されている), (2) は、初音ミクのライブコンサートで聴衆がサイリウムを振っている動画 (7 人の男性と 2 人の女性アバタ), (3) はサッカーのゴールシーンのオムニバス動画 (6 人の男性と 2 人の女性アバタ) である. 被験者である 20 代の男性 7 人と女性 1 人に上記 (1), (2), (3) それぞれのアバタの動きをアクティブなものから順位付けしてもらった. 評価の際、YouTube 動画は隠し、アバタはすべて男性として表示した.

アルゴリズムにより計算されたアクティブ度の順位と主観評価に基づいたアクティブ度の順位との関係を示すスピアマン順位相関係数を、アルゴリズムと動画の種類別にまとめたのが表 1 である.



図 10 評価に用いた YouTube 動画
Fig. 10 YouTube videos used in experiments.

表 1 アルゴリズムと主観評価順位のスピアマン順位相関係数
Table 1 Spearman's rank correlation coefficient between algorithms and subjective evaluation.

アルゴリズム	動画(1)	動画(2)	動画(3)
A	0.62	0.26	0.7
B	0.41	0.6	0.55
C	0.54	0.7	0.81

動画 (1) では、体操の進行にともない全体として視聴者であるアバタは全身を大きく動かしていた. 体操の特徴として、胴体の位置をあまり動かさずに、手・腕・足先・腿などの各部位を動かす運動が多いため、SpineBase と手や足などの骨格点との距離が大きく変動し、A の相関係数が高くなった. また変動の少ない骨格部位が多かったので、アルゴリズム B のアクティブ度が小さくなり、相関係数も低くなったと考えられる. 動画 (2) では、アバタはステージ上のパフォーマンスを観ながら、腕を振るだけで大きな動きはほとんどなく、アルゴリズム A の評価との相関係数が低くなった. 一方で、腕を振る動作が頻繁に起こるためフレーム間差分が大きくなり、アルゴリズム C の評価との相関係数が高くなった. 動画 (3) では、サッカーのゴールが決まったときに右や左に動き回ったりするなど大きなアクションをとるアバタがいたため、距離変化が大きくなり A と C の相関係数が高くなったと考えられる. これらをまとめると、SpineBase を基準とした骨格点との距離をもとにしたアルゴリズム A は体の各部位の動きをとらえるのに適しており、頻繁で速い動きをとらえるにはフレーム間差分をもとにしたアルゴリズム C が適している. 全体として動画の種類によらず、フレーム間差分によるアクティブ度は主観評価と高い相関があるといえる.

5.2 システムの使用に関する評価

20 代の男女 11 人に協力してもらい本システムについてアンケートに回答してもらった. 使用した動画は音楽 (図 11 左) と体操 (図 11 右) のビデオの 2 種類である. これらの動画を視聴しながら、動画に合わせてダンスや体操をするなど自由に動いてもらった. 手を用いて他のアバタを押しつける押しつけ機能も使用してもらった.

アンケートでは、被験者に設問に対する 5 段階 (5: とてもそう思う, 1: まったくそう思わない) で評価点をつけ理由を記入してもらった. 主な質問項目と結果を以下に示す.



図 11 評価に用いた WakWak Tube 動画

Fig. 11 Screenshots of WakWak Tube for evaluation.

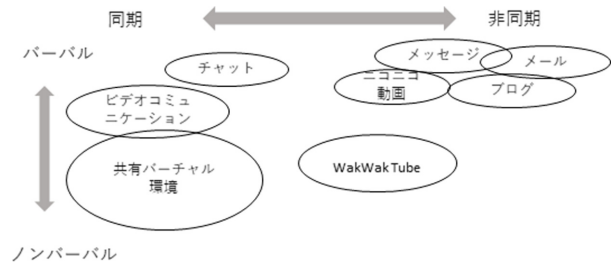


図 12 遠隔コミュニケーションツールの分類

Fig. 12 Map of remote communication tool.

(1) アバタを用いるシステムは面白かったか？ その理由

平均評価点：4.4

主な理由

- アバタの動きを見ているだけでも楽しめた。
- アバタを押すことができるので同じ場所にいるように感じた。
- ダンスの振り付けを知らないと一体感を得にくい。
- (図 11 左の動画において) 具体的に何をすればいいのかわからなかった。

(2) アバタを押しおける機能は面白かったか？ その理由

平均評価点：3.8

主な理由

- 押しおける機能によって自分のアバタを見やすくできるのが良かった。
- 普通に踊っていても (他アバタを) 押しおけてしまった。

図 11 左の動画では、歌手の踊りの振り付けを真似するのが難しかったため、どのような動作をしてよいか分からず戸惑いがうかがわれた。図 11 右の動画では、動画の体操の動きに従って動作すればよいことが明らかで、他のアバタも一緒に同じ動作をしているため、動きが自然に誘発された。押しおける機能や相手がよけてくれる機能により、ほとんどのユーザが他のアバタと同じ場所にいるように感じ、臨場感・一体感を醸成するのに有効であった。

6. 非同期コミュニケーションシステムとしての WakWak Tube

6.1 コミュニケーションシステムとしての WakWak Tube の位置づけ

コンピュータを介した遠隔コミュニケーションツールを、バーバル・ノンバーバル情報の軸と同期・非同期システムの軸で整理したのが図 12 である。

バーバル情報を扱うツールは、スカイプ (ビデオコミュニケーション) やツイッター (マイクロブログ) など商用サービスとして広く普及しており、SNS はこれらのツールが複合されたサービスが提供されていると見なすことができる。またセカンドライフ [13] のような共有バーチャル環境 (SVE: Shared Virtual Environment) を利用したコミュニケーションは、バーバル情報が中心であるが、アバタ表

示によりさまざまなノンバーバル情報も共有することが可能である。一方ノンバーバル情報に関する研究として、ビデオコミュニケーションにおける視線など、会話成立に重要なアウェアネス情報の伝達に重きを置いていた初期のグループウェア (たとえば ClearBoard [14] など) があげられる。また人の表情や身振り以外のノンバーバル情報を共有する研究では、呼吸や心拍を遠隔に伝えるデバイス Mobile Feelings [15] がある。このシステムは、パートナーの身体情報を共有し心理状態を推測することを意図している。さらに人そのものの情報以外に、環境や物の情報を伝達するシステムとして見守りポット [16] がある。ポットの利用状況を遠隔の家族に自動的に知らせることで、一人暮らしの高齢者の見守りの一助とするものである。ノンバーバル情報のモダリティは多岐にわたるため、今後もさまざまなデバイスや手法の提案が期待される。

一方、同期・非同期に眼を向けてみると、現在主流なのは、ツイッター、LINE、Facebook など用いられているブログやメッセージ交換アプリケーションなど非同期のツールである。ビデオコミュニケーションなどの遠隔同期型コミュニケーションは、遠隔で同期して参加するという時間的な制約が大きいに加え、状況のアウェアネス情報が不足するため、沈黙が気まずい雰囲気や冗長さなど心理的な圧迫も大きい。上記のようなメッセージ交換アプリケーションでは、相手のメッセージやコメントに瞬時に反応してメッセージを送りかえすような、非同期ではあるものの一連のコミュニケーションが連鎖するような (同期的な) 使い方が行われている [6]。

ニコニコ動画は、ユーザの動画視聴およびコメント投稿は非同期的に発生するが、コメント表示を動画再生のタイムラインと同期させることで、ユーザに他のユーザと「今、ここで」一緒に視聴しているかのようなライブ感を醸成させ、あたかも体験を共有しているような感覚を持たせている点に特徴がある。WakWak Tube も同様のアーキテクチャを持つ非同期型システムである。大きく異なるのは、ニコニコ動画がテキストによるコメントというバーバル情報であるのに対して、WakWak Tube は身体情報というノンバーバル情報を扱う点にあるが、以下 WakWak Tube の特徴を詳細に見ていく。

6.2 WakWak Tube のアーキテクチャ

WakWak Tube は動画視聴をベースにした身体情報を共有するために、劇場メタファを採用している。ステージとなる動画表示領域の下に、客席に見立てた視聴者領域が設けられて、視聴者のアバタが表示される。この共有空間には以下の特徴がある。

(1) 他者の存在と矛盾がない自己投射性

視聴者領域の中心に自分自身のアバタが配置され、リアルタイムに体の動きが反映されることで、バーチャル環境に投射性を持った自己プレゼンスが実現される。自身のアバタと他者の過去アバタは、同じ動画を視聴しているというコンテキストが一致しており、動作の時間軸は動画再生のタイムラインに一致しているので、同時に表示されたときに違和感が生じることがなく、「今、自分があたかも他の人と一緒に」という同時存在の感覚を引き起こすことができると考えられる。ニコニコ動画のようなコメントの表示では、このような自己投射性の実現は難しく、身体情報を扱うことによって可能になると考えられる。

(2) ユーザによる空間の再構成

アバタの表示領域が限られているため、ユーザはアバタセレクションにより表示するアバタを選択する。WakWak Tube で実装されている選択基準は、アバタが記録された日時と、アバタの性別・アクティブ度であるが、よりさまざまな属性を持たせて選択項目に付け加えることができる。またユーザ自身が移動したり、押しのけや引き寄せといった機能で他アバタを移動させたりして、空間を再構成することも可能である。またこの空間はユーザごとに異なったアバタで構成されるために、他のユーザとビューが共有されていないことが特徴である。

(3) コンテキストと時間軸の一致を利用したインタラクション

上で述べたコンテキストと時間軸が一致するという特性のもとで、非同期に蓄積されたアバタに対しても押しのけたり引き寄せたりといった位置変更のインタラクションが成立する。このようなインタラクションはバーバル情報のみを共有するニコニコ動画では実現できない。インタラクションにより「今、あたかも自分の意志で」行っているかのような感覚を生じさせることができると考えられる。また身体的なインタラクションはバーチャル環境でのさまざまな体験を生じさせ、6.3 節で述べるようにユーザ体験の共有にも拡張することができる。

6.3 WakWak Tube のコミュニケーション機能の拡張

(1) バーバル情報

ニコニコ動画では、文字情報によるコメントの書き込み・再生メカニズムを提供しているが、WakWak Tube では、ユーザ音声録音しておき、動画再生時に音声も同時再生することにより、臨場感を向上させることが考えられ

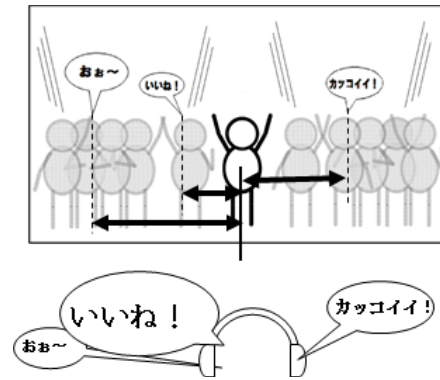


図 13 音声再生

Fig. 13 Voice playback.

ようこそ Aさん !!
この人たちがあなたを選びました
J F U X R D
この人たちがあなたに「いいね」しました
R C M
この人たちがあなたにメッセージを送りました
D B

図 14 メッセージ通知パネルの例

Fig. 14 Example of notification panel.

る。たとえばコンサート会場では、ステージのパフォーマンスだけでなく、周囲の歓声や拍手なども「今、ここで」ライブ演奏を聴いているという感覚に寄与している。異なる時間に録音された音声を再生するに際して、自己を中心に左右にアバタが配置されているので、図 13 に示すように、距離が近い人ほどボリュームを大きくするなど、空間構成に応じた音声再生が必要と考えられる。

(2) WakWak Tube でのユーザ体験の共有

コンピュータでの動画鑑賞やゲームプレイなどバーチャル環境での生活時間が長くなり、ゲームプレイを動画でアップロードして共有するといったように、バーチャル環境における体験を共有したいという欲求が生まれてくるのは自然な流れである。ニコニコ動画では投稿されるコメントにより動画視聴しているという体験が共有されるが、WakWak Tube では他アバタとの身体情報が共有されインタラクションが生じることで、より多くの体験を引き起こすことができる。たとえば、引寄せ機能を利用して自己アバタの隣に引き寄せたこと、また踊りが上手なアバタに「いいね」をマークすることなどである。これらの事実を動作の対象となったアバタユーザに、図 14 のようなパネ

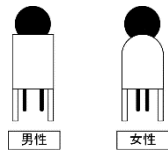


図 15 着席アバタの例

Fig. 15 Example of sitting avatar.

ルを用いて通知することができる。このようにバーチャル環境での自己体験をメッセージとして他のユーザに知らせ共有することは、さらなるコミュニケーションのきっかけになることが期待できる。

これまで述べてきた例では、ユーザは WakWak Tube システムの前に立って体を動かしながら動画を視聴することを想定してきたが、多くの場合椅子に腰掛けたり、寝転がったりしながら動画を楽しむことが多い。このような場合でも他のアバタとの一体感を得るために自己アバタの表現を検討する必要がある。たとえばあまり動作しないユーザや動作表示を必要と感じないユーザの場合、図 15 のような表示を用いるのも一案である。

7. おわりに

オンデマンドの動画サービスの視聴者の非同期コミュニケーションを実現するために、ユーザの身体動作というノンバーバル情報に注目し、WakWak Tube を開発した。このシステムの特徴として、

- (1) 自己投射性と非同期の他のユーザとの同時存在性
- (2) ユーザによる空間再構成とユーザ独自のビュー
- (3) 動画再生時間軸での他アバタへのインタラクションがあげられる。さらに機能拡張により
- (4) バーバル（音声）コミュニケーション
- (5) ユーザ体験の共有

の可能性について述べた。今後 WakWak Tube を、動画視聴をベースとした非同期コミュニケーションの機能拡張に耐えうるようなプラットフォームとして充実させ、実用的な機能実現のための開発につなげていく予定である。

謝辞 WakWak Tube の開発に携わってくれた若月友祐君に、感謝の意を表す。

参考文献

[1] YouTube, available from (<https://www.youtube.com/>) (accessed 2018-01-02).

[2] Oehlberg, L., Ducheneaut, N., Thornton, J.D., Moore, R.J. and Nickell, E.: Social TV: Designing for distributed, sociable television viewing, *Proc. European Conference on Interactive TV '06*, pp.251-259 (2006).

[3] Chorianoopoulos, K.: Content-enriched communication—Supporting the social uses of TV, *Journal of the Communications Network*, Vol.6, Part 1, pp.23-30 (2007).

[4] 渋谷明子, 志岐裕子, 李 光鎬ほか: SNS 利用者のコミュニケーションとテレビ視聴, *メディア・コミュニケーション*, No.62, pp.57-78 (2012).

[5] niconico, available from (<http://www.nicovideo.jp/>) (accessed 2018-01-02).

[6] 濱野智史: アーキテクチャの生態系—情報環境はいかに設計されてきたか, *ちくま文庫* (2015).

[7] Coppens, T., Vanparijs, F. and Handekyn, K.: Amigo-TV: A social TV experience through triple-play convergence (White paper), Murray Hill, NJ: Alcatel-Lucent (2005).

[8] Harrison, C. and Amento, B.: CollaboraTV: Using asynchronous communication to make TV social again, *Adjunct Proc. EuroTV*, pp.218-222 (2007).

[9] 代蔵 巧, 棟方 渚, 小野哲雄: ExciTube: 鑑賞者の興奮を共有する動画鑑賞システム, *日本バーチャルリアリティ学会論文誌*, Vol.18, No.3, pp.247-254 (2013).

[10] 吉田有花, 宮下芳明: 身体動作の重畳表示による動画上での一体感共有, *情報処理学会インタラクティブ 2012 論文集*, pp.527-532 (2012).

[11] 川井康寛, 志築文太郎, 田中二郎: 動画共有に基づいた非同期コミュニケーションの一体感を向上させるインタフェース, *情報処理学会第 70 回全国大会予稿集*, pp.1-715-716 (2008).

[12] Vatavu, R.: Audience silhouettes: Peripheral awareness of synchronous audience kinesics for social television, *Proc. TVX '15*, pp.13-22 (2015).

[13] Second Life, available from (<http://www.secondlife.com/>) (accessed 2018-01-14).

[14] Ishii, H. and Kobayashi, M.: ClearBoard: A seamless medium for shared drawing and conversation with eye contact, *ACM Proc. CHI '92*, pp.525-532 (1992).

[15] Sommerer, C. and Mignonneau, L.: Mobile Feelings: Wireless communication of heartbeat and breath for mobile art, *Proc. ICAT2004*, pp.346-349 (2004).

[16] みまもりほっとライン, 入手先 (<http://www.mimamori.net/>) (参照 2018-01-03).



坂内 祐一 (正会員)

神奈川工科大学. 1980年早稲田大学大学院理工学研究科修士課程修了, 1988年ミシガン州立大学大学院コンピュータサイエンス学科修士課程修了, 2007年慶應義塾大学大学院理工学研究科博士後期課程修了. キヤノン(株)情報メディア研究所画像メディア研究室長, 千葉大学特別研究員等を経て, 2011年より神奈川工科大学情報メディア学科教授. 2007年論文賞受賞. 人工・拡張現実感, 五感情報処理等に関する研究に従事. ACM, IEEE-CS, 日本 VR 学会, 人工知能学会, 味と匂い学会各会員, 博士(工学).