「マルチメディア，分散，協調とモバイル
(DICOMO2018)シンポジウム」 平成30年7月

動画視聴における他視聴者の音声情報の再生方式

松長雄也[†1]　谷中俊介[†1]　坂内祐一[†1]

概要： YouTube などの動画共有サービス上におけるユーザ間のコミュニケーション手法として，身体動作やインタラクション機能に加え，音声情報を扱うことが可能なシステム WakWak Tube を開発した．このシステムでは，動画を視聴しているユーザの身体動作を Kinect によって取得，記録し，アバタと呼ばれる 2D のグラフィックで表示することが可能である．ユーザはアバタを介して他視聴者とコミュニケーションを取ることができる．本研究では，従来の WakWak Tube では扱えなかった音声情報を扱えるようにした上で，本システムにおいて音声を用いることの有用性を調査する評価実験を行った．結果，アバタと音声の対応付けを行うことで，音声が効果的に作用していることが示唆された．

Reproduction Method of The Sound Information of Other Viewers in Video Viewing

YUYA MATSUNAGA[†1]　SHUNSUKE YANAKA[†1]　YUICHI BANNAI[†1]

1. はじめに

YouTube やニコニコ動画などの動画共有サービスが登場してから 10 年以上もの年月を経た今日であるが，現在も様々な動画がアップロードされ続けている．近年では，“YouTuber”などといった呼称も生まれ，より注目を集めている．こうした動画共有サービスの多くには，動画を視聴したユーザによってテキストベースのコメントを投稿でき，ユーザ間でのコミュニケーションを促すようなシステムが導入されている．

そうした中，TV や動画共有サービス上におけるコミュニケーション手法に関する研究が数多くなされている[1]．吉田ら[2]は動画を視聴しているユーザの身体動作を動画上に重畳表示し，非言語によるコミュニケーションを通じて一体感の向上を試みている．高野ら[3]はストリーミング動画においてテキストコメント及び音声会話によるコミュニケーション手法を提案している．若月ら[4]は非同期の同一動画視聴者を対象としたコミュニケーションツールとして WakWak Tube を開発している．このシステムは動画を視聴しているユーザの身体動作を取得し，アバタという 2D のグラフィックにその身体動作を反映させ，動画と共に表示している．アバタとして記録された身体動作はその時に再生された動画に紐付けて保存されており，再度同一の動画が再生されると過去アバタとして表示される．これより，アバタはリアルタイムに動画を視聴しているユーザのものと，過去に同一の動画を視聴した他視聴者のものが表示される仕様となっている．更にユーザはアバタを介して過去のアバタ(以下：過去アバタと呼ぶ)に触れ，押しのけるなどのインタラクションを起こすことも可能である．これらの機能を通じて非同期の他視聴者とコミュニケーションを行える場を提供することで，一体感や臨場感を得られるシステムを提案している．

テキスト情報なのか，ジェスチャーなどの身体動作情報なのかに拘わらず，一体感や臨場感を得るにはユーザやアバタの感情表現及び，他視聴者がそれをどのように受け取るかが大きな要因になると考える．他人の感情を判断する場合，特に相手の表情を伺うことが不可能な状況において，音声は多大な役割を果たすとされている[5]．従来の WakWak Tube では，アバタを介した身体動作による自己プレゼンス及び，インタラクション機能によってコミュニケーションを図っていた．しかし，それだけでは相手の感情を判断するには十分とは言えない．

そこで本研究では，WakWak Tube にユーザの発した音声情報を扱える機能を追加し，アバタと対応付けることによって，身体動作とインタラクション機能及び，音声情報という 3 つの要素によってユーザの一体感や臨場感の更なる向上を試みる．そして本稿では，新たに実装した音声処理機能に対する評価実験を行った．

2. 提案手法

若月らの開発した従来の WakWak Tube に対して，新たに視聴者の音声情報を扱う処理を加えることで，他視聴者との一体感や臨場感の向上を目的とした新たな WakWak Tube を開発した．本システムを用いた動画視聴ページの画面を図 1 に示す．WakWak Tube は，ユーザをリアルタイムに表示すると同時に記録も行うことで，他視聴者が同一の動画

†1 神奈川工科大学情報学部情報メディア学科

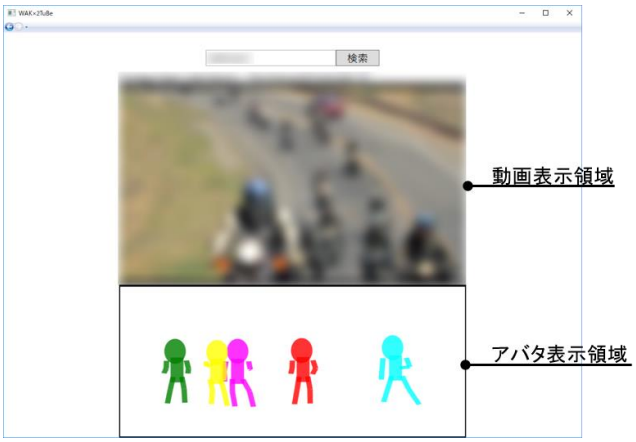

図 1　WakWak Tube における動画視聴ページの画面

を再生した際に過去アバタとして表示することが可能となっている．画面上部には選択した動画が表示される動画表示領域があり，画面下部にはアバタが表示されるアバタ表示領域がある．アバタ表示領域にて中央に表示されている赤色のアバタがリアルタイムのユーザのアバタ(以下：ユーザアバタと呼ぶ)であり，その他の周りに表示されているアバタが過去アバタである．ユーザはこのアバタ表示領域内で動き回ることや，過去アバタにインタラクトすることによって他視聴者とコミュニケーションを取ることが可能である．

2.1　システム構成

WakWak Tube のメインプログラムは C#にて構成されており，またアバタと動画を同期的に再生させるため，JavaScript による動画制御が行われている．視聴者の骨格情報の取得には，Microsoft 社から販売されている Xbox One Kinect センサ(以下：Kinect V2)を用いている．ユーザの音声の取得には NAudio を用いている，詳細は 3.1 節にて後述する．

2.2　音声の取得とアバタ間距離に応じた音量調整

本稿では，アバタ間距離に応じた音量調整機能を提案する．提案する手法によって，ユーザの骨格情報に加え音声情報の取得も行えるようにし，アバタと共に提示することによって，より一体感や臨場感を感じられるようなシステムを目指す．他視聴者の音声の提示方法として今回は，アバタ表示領域内におけるアバタ同士の位置関係によって，ユーザに聞こえる過去アバタの音声に変化を持たせるという手法を提案する．アバタ表示領域内においてリアルタイムのユーザアバタと過去アバタの間が近距離であるほどその過去アバタから聞こえる音声の音量が増加され，逆に間が遠距離であるほど音量が減少される．これにより，感情表現に重要な役割を果たすとされる音声の提示，それによる一体感の向上及びアバタ表示領域内のリアリティが増すことによる臨場感の向上が見込まれるのではないかと考えた．

従来の WakWak Tube でも，ユーザはアバタ表示領域内で自由に動き回れ，他アバタを押しのけることも可能であったが，それらの行動を起こすための明確な理由は存在しなかった．しかし，アバタに音声情報を付与し，アバタ同士の位置関係を対応付けることによって，ある音声を発しているアバタに近づいたり，あるいは押しのけて遠ざけたりなど，アバタ表示の機能を活かしつつ，それらの行動を起こす動機付けのためにも，音声情報の追加は大きな役割を果たせるのではないかと考えている．

3.　システム構築

3.1　音声情報の取得

音声情報の扱いには NAudio という.NET 用の音声処理ライブラリを用いている．録音の開始及び停止は動画の再生及び終了を基準としており，これらはアバタの再生や記録と同期的に行っている．音声の録音にはコンデンサーマイクやマイク付きヘッドセットなど，何らかの外部録音機器を用いることを想定しており，本稿では主にコンデンサーマイクによって音声の録音を行った．

3.2　アバタ間距離の算出

アバタ間距離に応じた音量の増減を実現するためには，アバタ表示領域内の座標系におけるリアルタイムのユーザアバタ座標及び，過去アバタ座標の取得が必要である．そこで本システムでは，Kinect V2 にて取得可能な骨格座標である，ユーザの頭の位置を認識する Head の関節点を用いて，各アバタの位置を示す座標としている．Head の関節点を用いた理由としては，人間が声を発する口や，音を聞く耳はいずれも頭に位置していることから，本システムではアバタの位置を示す座標に Head の関節点を用いた．

Kinect V2 を用いた骨格情報の取得では，3 次元空間における座標値で計測しているが，アバタ表示領域は 2 次元平面であるため，取得した座標値は 2 次元座標に変換する必要がある．アバタ間距離の算出方法の概要を図 2 に示す．アバタ間距離の算出方法は過去アバタの Head の X 座標値からユーザアバタの Head の X 座標値を減算させ，更に得られた値を絶対値に変換するといった方法である．図 2 では，A と C が過去アバタ，B がユーザアバタとなっており，各アバタの下部にはアバタ表示領域におけるそれぞれの Head の X 座標値を例として示している．図 2 のようなアバタ配置において，過去アバタ A からユーザアバタ B の X 座標値を減算する場合のような，過去アバタの X 座標値がユーザアバタの X 座標値よりも小さい場合，得られる値が負になってしまい距離の値としては適していない．そこで，その値を絶対値に変換することで，過去アバタがユーザアバタの左右どちらに位置しているかに関わらず，正の値を

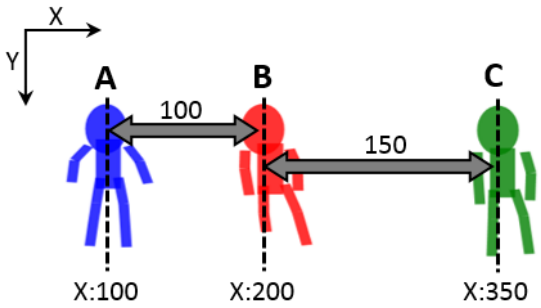

図 2　アバタ間距離算出の概要

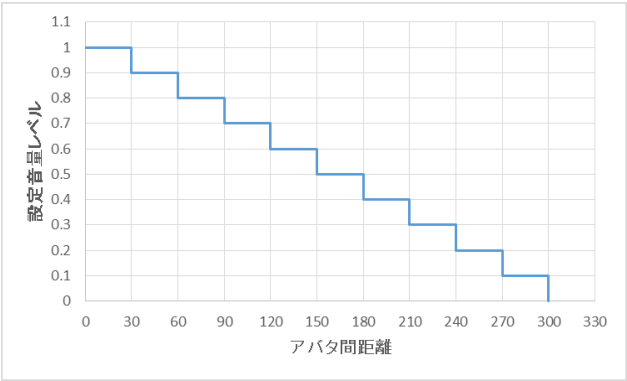

図 3　アバタ間距離による音量の増減傾向

返すような処理を加えている．また，算出した数値は音声処理時の計算を簡易化する目的で，小数点以下の値は切り上げている．

3.3 算出距離に基づいた音量調整

本節では，アバタ間距離に基づいた音量設定を行う際に考慮した各種条件及び，それに伴う音量設定方法について述べる．まず，本研究での WakWak Tube の運用条件について説明する．WakWak Tube ではユーザアバタ 1 体以外に最大 8 体までの過去アバタを同時に再生させることが可能である．しかし，現状のシステムではアバタ表示領域内に 8 体もの過去アバタが存在した場合，音声の識別が困難になることが予想された．そこで，本研究ではユーザアバタ 1 体，過去アバタは最大 4 体までの運用を想定して開発を行っている．

次に，過去アバタの配置に関する条件である．1 体目は X 座標 100 の位置，2 体目は X 座標 150 の位置，といったように過去アバタには表示される際の初期位置座標がそれぞれ設定されている．本システムでは，この初期位置座標を考慮して，音量が最大値となるアバタ間距離及び，最小値となるアバタ間距離などを設定している．最後に，音量の増減幅についてである．NAudio にて設定可能な音量レベルは float 型で 0.0f～1.0f であるため，この数値内で音量をコントロールする必要がある．

以上の条件より設定した音量の増減傾向を図 3 に示す．音量レベルは 0.0f～1.0f の 11 段階に分け，アバタ間距離が 30 毎に増減することで音量レベルもそれに対応して増減する．アバタ間距離が 0，つまりユーザアバタと過去アバタがほぼ接触しているような状態では，音量は最大値の 1.0f にて再生され，アバタ間距離が 300 以上離れている場合には音量は最小値の 0.0f，つまりその過去アバタからの音声が聞こえなくなるという設定にした．

ここで，音量が最小値になるアバタ間距離を 300 以上と設定した理由について述べる．これには，前述した過去アバタの初期位置座標が関係している．本稿における WakWak Tube の運用想定では，過去アバタは最大 4 体までとしている．仮に過去アバタを 4 体表示する場合，ユーザアバタを中心に，左右に 2 体ずつ配置される．その際に，ユーザアバタからより離れた位置に配置される過去アバタが左右 1 体ずつ生じる．動画視聴開始時，ユーザにその過去アバタの音声が聞こえるか聞こえないか程度の音量設定を行うことで，ユーザがその音声に興味を持てば，近づくなどの行動を起こす可能性がある．そうした動機を引き起こす要因として，聞こえるか聞こえないか程度の音量設定が適していると考えた．その際の過去アバタとユーザアバタのアバタ間距離が約 300 であったことから，今回は音量レベルが最小値となるアバタ間距離を 300 以上と設定した．

4. 評価実験

男女 6 名（21.8±0.7 歳，mean±S.D.）を被験者として，アバタ間距離に基づく音量増減の処理に関する評価実験を行った．本研究での最終的な目的は音声情報を取り入れることによって，一体感や臨場感が向上するか調査ことである．そこで本稿ではまず，ユーザが本システムを通じて音声変化を感じることが可能なのか，またはアバタと音声に対応関係を見出せるのか調査を行った．

本稿では実験 1～3 を実施した．実験 1 は過去アバタが 1 体，音声が 1 つ．実験 2 は過去アバタが 3 体，音声が 2 つ．実験 3 は過去アバタが 2 体，音声が 2 つという条件である．実験 1 は音声変化を感じられるか確かめること，実験 2 と実験 3 はアバタと音声の対応関係を被験者が認識に，正確に判別できるか否かを確認することを目的とする．いずれの条件においても実験環境は図 4 のような構成で行った．Kinect V2 は PC モニタの上部に設置されており，被験者は Kinect V2 から約 1.7m 後ろに離れた位置に立たせた．

また，本実験で表示する過去アバタは全て実験者が実験の意図に沿うように事前に記録したものであり，音声に関しては今回の実験目的を考慮し，インターネット上で無料配布されているナレーション音声を用いている．音声ファイルは，音楽編集ソフトウェア Audacity にて正規化し，音量を統一している．更に，今回の実験では動画の内容は考慮しないため，実験時間を計測するものとして 10 秒をカウントダウンする動画を使用している．

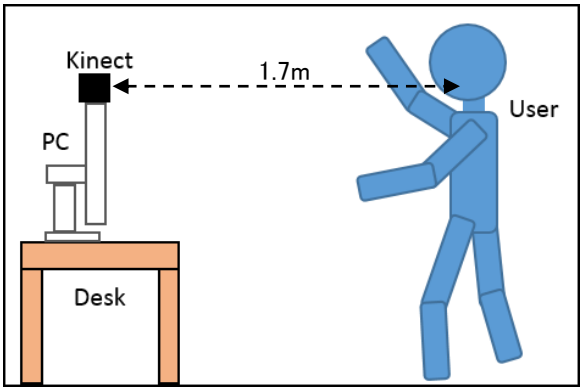

図 4　実験環境概要

4.1 実験 1

実験 1 では, ユーザアバタと過去アバタが 1 体表示され, その過去アバタが音声を発している．被験者にはアバタ表示領域内を自由に動き回ることを許可した．実験終了後に実施したアンケートでは，過去アバタの音量が変化していると感じたか(質問：Q.1-1)，音量増減によって過去アバタとの距離感に変化を感じたか(質問：Q.1-2)，の 2 点について 5 段階評価で回答させた．

4.2 実験 2

実験 2 では，ユーザアバタの右側に 1 体，左側に 2 体の過去アバタが表示され，左側 2 体が音声を発している．更に左側の 2 体は 10 秒の間，徐々にお互いの位置が入れ替わっていく．この時，被験者には画面中央からの移動を禁止している．実験終了後に実施したアンケートでは，左側 2 体の過去アバタの位置が変化しているように感じたか(質問：Q.2)，5 段階評価で回答させた．

4.3 実験 3

実験 3 では，ユーザアバタの左右に 1 体ずつ過去アバタが表示され，左側の過去アバタの音声は男性の声，右側の過去アバタの音声は女性の声を発している．また，右側の過去アバタは短い距離を左右に往復し続けている．被験者には左側の過去アバタが男性と女性のどちらの声を発していたか(質問：Q.3)，実験終了後に実施したアンケートにて回答させた．

5. 結果，考察

実験 1 と実験 2 のアンケート結果を表 4 に，実験 3 のアンケート結果を表 5 に示す．本実験の結果より，本稿にて提案したアバタ間距離に基づく音量増減処理において，音声の変化を感じられていること及び，それに伴う，アバタと音声の対応化が行えていることを確認した．更に実験 3 では被験者 6 人全員が正確に判別できていた．これらの結果より, 音声情報が WakWak Tube において有用性のある要素だと示された．

表 4　実験 1 と実験 2 のアンケート結果(回答数)

実験番号	実験 1		実験 2
質問番号	Q.1-1	Q.1-2	Q.2
全くそう感じなかった	0	0	0
そう感じなかった	0	0	0
どちらとも言えない	0	0	1
そう感じた	0	2	3
とてもそう感じた	6	4	2

表 5　実験 3 のアンケート結果(回答数)

実験番号	実験 3
質問番号	Q.3
男声	0
おそらく男声	0
分からない	0
おそらく女声	0
女声	6

また，質問 2 にて回答に多少のバラつきが見られたが，被験者から得られた意見として, 実験 2 では 10 秒という短い時間の間にアバタの入れ替わりが行われたため，音の増減は分かるもののアバタの位置が変化しているようには感じ難かったとしていた．したがって，実験内容によっては実験時間の調整及び，動画選択の重要性が示唆された．

6. おわりに

本稿では，アバタ間距離に基づいて音声の音量増減が可能は WakWak Tube を開発した．これにより，身体動作とインタラクション機能に加え，音声情報によるコミュニケーションが図れるようになった．今後はモノラル再生されている音声をステレオ再生に対応させ，左右のアバタ位置に応じてヘッドセットから聞こえる音声を調整する機能や，動画コンテンツに付随して適当な音声再生を行う機能などを検討し，音声情報による一体感や臨場感の更なる向上を目指していく．

参考文献

[1] C.Harrison and B.Amento,CollaboraTV:Using asynchronous communication to make TV social again,Adjunct Proceedings of EuroTV,pp.218-222(2007).
[2] 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ション 2012 論文集,pp.527-532(2012).
[3] 高野祐太郎,田島孝治,大島浩太,寺田松昭:投稿型動画聴におけるユーザ間リアルタイムコミュニケーション支援システムの提案,電子情報通信学会論文誌 D Vol.J93-D No.10,pp.2302-2316(2010).
[4] Yuichi Bannai,Yusuke Wakatsuki:WakWak Tube:An Asynchronous YouTube Co-Viewing System,Proc.International Workshop on Informatics(IWIN2017),pp.159-164,(2017).
[5] Micheal W Kraus:Voice-only communication enhances empathic accuracy,American Psychologist,pp.644-654(2017).