

異種環境での感情表出アバターを用いたリモート対話環境

宇都木契¹ 藤原貴之¹ 栗原恒弥^{†1}

概要: 遠隔営業や説明議論の支援を目的としたコミュニケーション支援システムの試作を行なった。提案システムは、(1)HMD,PC,スマートフォンなど異種機器を対象として Web ブラウザ上に対話・議論用の仮想空間を構築する (2) 顔部分を撮影するカメラからユーザの表情リアクションを取得し、空間上のアバター動作に反映する(3) 視聴コンテンツの共同再生機構と表情ログのロギングを合わせ会話の状況分析を行うためにリアクションKPIの変化を計測する機構の三点を特徴とする。本発表では、この試作で検討した会話に対して顔表情による操作に基づく非言語情報のフィードバックを行う機構の実現と、同システムに多様な接続機器で接続した場合のユーザビリティを担保するための画面構成デザインについての議論を紹介する。特に、異種機器の環境においてディスプレイの面積に基づく行動差異が発生することを前提として、共有ドキュメントの視聴行動を複数の参照モードに分けて管理、視聴共有と部分注目、逸脱参照といった状態を明示的に区別して管理する手法について論じる。

Remote Communication with Interactive Emotional Feedback in Asymmetrical Environments.

KEI UTSUGI¹ TAKAYUKI FUJIWARA¹ TSUNEYA KURIHARA¹

1. 背景

COVID-19 対策の影響で、各種の社会活動が遠隔コミュニケーションでの代替が進むにつれて、リモート化において失われた非言語コミュニケーションの重要性の論議も活発化している。典型的な対策としてはカメラ映像で顔表情を提示する TV 会議方法が用いられる。が、その一方、顔や室内といったプライベート情報を映像で開示する際のセキュリティリスクから、特に不特定者が混ざる会合ではカメラ画像の開示が敬遠されることも多い。

これに代わる新たな参画方式として、VR・メタバース形態でのコミュニケーション技術に注目が集まりつつある。メタバース形態のコミュニケーションでは、本人の代替物であるアバターを用い、ジェスチャやキャラクター的認知などの非言語コミュニケーションを表現する[1]。匿名性が担保された商空間の構築や、各種マルチモーダル分析技術の適用など、大きく期待されている分野となっている。

だが現状のビジネスシーンにはメタバースの導入をとどめている要因がいくつかある。その一つが機器形態の問題である。多くのメタバースシステムは HMD や高性能 PC などを含む高級環境を想定しているが、現在のところ、ビジネスでのリモートワーク用途ではノート PC、個人のリモートコミュニケーション用途ではスマートフォン等が用いられることが多い。どちらの機器もディスプレイ表示領域には制約があるため、共有文書提示と非言語情報の演出はディスプレイの面積を奪い合う関係にあり、文書表示効率の邪魔とされて効果が曖昧なアバター表現が敬遠・排除さ

れてしまうことも少なくない。また、自機器に新規のソフトウェア環境をインストールすることにも抵抗がみられるなどの要因もある。

現時点で用いられている各機器環境に対してメタバース的なコミュニケーションを導入するためには、画面デザインとソフトウェアデザインの設計が大きな鍵となる。

本研究では、PC やタブレット、スマートフォンなどの一般的な環境で、簡便に VR 的なコミュニケーションが行える環境の試作を行った。以下の三点を特徴としている。

(1) JavaScript ベースでシステムを構成し、HMD,PC,スマートフォンなど様々な機器の Web ブラウザ上で動作する。

(2) 感情表出機構の搭載: Web カメラの顔画像から、表情判別とジェスチャ判別を行い、利用者アバターの操作や話の内容への反応を表出する。

(3) 視聴コンテンツの共同再生: メタバース型 3D 空間と 2D 画面の双方で文書や動画などの各種メディアを携帯・PC など多様な再生環境で共同視聴しながら対話を行う。

本発表では、これらの試作から得られた知見について報告する。第 2 章では先行研究と論点の整理を行い、第 3 章でシステムの概要とユースケースを説明、第 4 章では顔表情からの非言語コミュニケーション抽出モジュールの説明、第 5 章では共同視聴行動と画面デザインについてそれぞれ開設を行う。第 6 章では感情認識モジュールの挙動検証評価を記載し、第 7 章において総括を述べる。

¹ 日立製作所研究開発グループ
Hitachi Ltd.

2. 先行研究

2.1 ビデオ会議のデザイン

ビデオ会議システムやアバターチャットシステムとして多くの機器やソフトウェア実装が提案されている。Meta社などはVR空間上にリモートデスクトップ環境を提示する例[2]を示しており、またMicrosoftは既存のビデオ会議システムであるTeamsと3Dアバター空間を連結させる構想について発表しており、ビデオ会議システムの顔表示領域にアバターの顔が出るなどのデザイン案が開示されている[3]。

2.2 表情の表出

本システムでは、互いの表情反応をリアルタイムに表出しコミュニケーションを促進する行為を取り扱う。表出と無意識の感情反応の検出は区別すべき点である。P.Ekmanは、本人の意図と無関係に微小動作として現れてしまう無意識の反応について述べている[4]が、多くのコミュニケーションにおいて、互いが隠そうとする本心の探り合いのような行動を避けるのがマナーであり、本システムではこの種の本人の表出意図と異なる無意識挙動は対象としない。

相手の発言に対し意図的に表出される行動の例として、頷き、首振りなどのジェスチャー動作や、賛同を示す笑顔などの表出がある。表情反応が文化に依存しないのに対して、これら意図的表出の使い方は文化モデルに依存する[4]。また相手に伝えようという意思に沿って表情も意図的に大きめに提示される傾向がある。また、ポジティブな表出行為をあえて誇張して行う事に、本人自身の意識へのフィードバック効果があることもしばしば議論されている[5][6]。

2.3 遠隔会話での相槌行為

相槌が会話を促進することについては発声での相槌、ジェスチャーの双方で分析が行われている[7][8]。だが遠隔会議システムでは、音声ノイズの自動削除や、多人数への配慮があり発声での相槌が行いにくいいため、顔表情の遠隔表出インタラクションの研究がされている[9][10]。

2.4 コンテンツの共同視聴

同一コンテンツを共に見て反応を共有する行為は、共同体の確立に重要な役割を示してきた。近年では、動画へのコメント共有等を中心に技術が発達している[11][12]。メタバース型のチャットは行動で非言語メッセージ反応を行うことが当初の目的であるが、ジェスチャーだけでは明確な反応を示すことも難しく、言語型のSNS同様により分かりやすくシンボルを提示する (e.g. 「いいね」アイコン、顔文字等)、音を鳴らすなど様々な反応共有システムが併用されていることが多い。

3. ユースケースとシステム概要

提案システムの構成を図1に示す。このシステムは、2名以上の複数ユーザの情報共有と対話に用いられる。

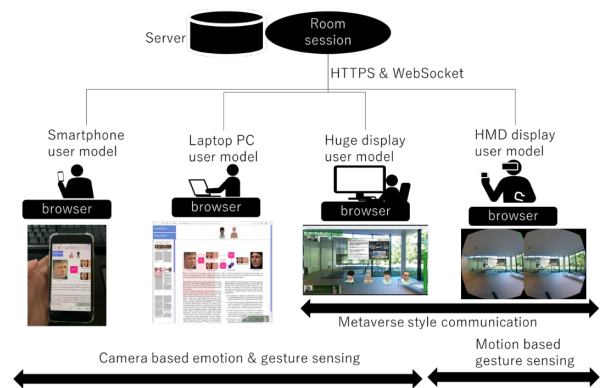


図 1 システム構成

Fig. 1. User diagram of target system.

システムは Web ブラウザに置かれたクライアント側とサーバ側で構成される。一つの会話セッションには複数の人数が参画し、ユーザの役割や所有する機器構成によって異なる画面形態で参加する。機器ごとの多様性やルームの再構築に対応させるため、各機能はモジュールごとに分解し、異なるシーンに組み合わせて再利用する。



図 2. VR ルーム型の運用事例

Fig 2. VR-room style integration

図2はセッションを3Dシーンとして表示する事例である。Webブラウザ上の3D表示にはBabylon.js[14]を使用している。3Dシーン内には話題を提示するコンテンツと共に、聴衆の顔反応を示すアバターが表示される。これらを動作させるための処理は、カメラを用いた表情分析モジュール(4章)と、各種のコンテンツ提示モジュール(5章)によって構成されており、各クライアントモジュールはWebSocketサーバを介して連携動作する。

4. 表情認識モジュール

本章では顔カメラの情報をコミュニケーションに用いるための機能を説明する。

本システムでは、顔カメラ情報は正確な3Dモーションキャプチャ再現のために直接用いるのではなく、コミュニケーションに必要な意図表出を抽出した結果を用いる。この抽出をロバスト化するため、顔アクションをKPI定量値に変換し、いくつかの特徴的ジェスチャーに分類・抽出する

処理を行っている。表情情報は、内心の心理状態ではなく、コミュニケーション目的に即して意図的に表出された大きな表情・ジェスチャを対象とし、意図的な表出の範囲外であろうネガティブな心情反応については原則として抑制的に用いる。

以下、実装形態(4.1 節)、KPI 設計(4.2 節)、アバター表現(4.3 節)についてそれぞれ記載する。

4.1 表情反応の取得モジュール

顔画像認識に基づくコントローラ部のモジュール動作例を図 3 に挙げる。モジュールには、顔画像取得と表情リアクションの判定機能、画像の WebRTC による共有機能、音声認識によるテキスト入力など、ユーザ入力に関するいくつかの機能が JavaScript モジュールとして纏められている。

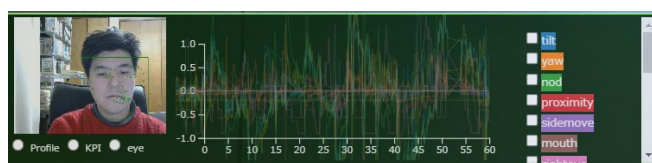


図 3 顔認識モジュール。顔画像からの特徴点認識と、各 KPI の時間遷移グラフ

Fig. 3 Face recognition module and emotion KPI visualizer

顔画像からの表情とリアクションを取得する処理には、OSS の faceapi.js[15]を使用している。faceapi.js は顔画像認識処理を行うライブラリで、顔特徴点の画像内位置認識と、感情分類の機能を持つ。サーバに画像を転送することなくクライアントブラウザ上の JavaScript 上で全処理が行われるため、顔画像の送信を行わずユーザプライバシーを担保することができる。

4.2 Web カメラ顔画像からの KPI 抽出

入力となる指標は faceapi.js から得られる顔特徴点の位置座標と、同様に得られる表情分類の情報である。

各指標についての処理の流れを図 4 に示す。

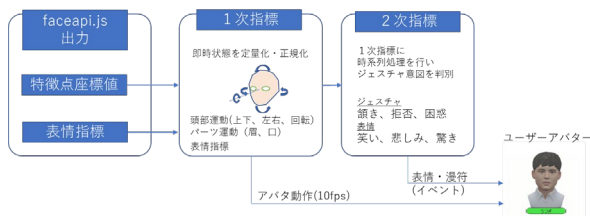


図 4 表情情報の指標化

Fig. 4 Generating feature KPI for facial expression

画像の二次元特徴量から、頭部と顔パーツの運動に相当する情報を抽出する 1 次指標と、それらを元に表出意図に関

連するジェスチャを抽出した 2 次指標に分けて処理される。1 次指標は 10fps のフレームレートで常時送信されてアバターのリアルタイム操作処理にも用いられる。2 次指標は表出イベントとして、アバター表示への漫符表現(4.3.2 節)や、コンテンツへの意見カウンティング(5.3 節)に用いられる。

4.2.1 1 次指標 (KPI 抽出)

第一段階処理は、いくつかの特徴点座標の移動量の組み合わせから、特定の頭部パーツの運動を示す独立パターン(頭部の運動や、眉、目、口の形態変化等)を取得する。この第一段階処理では、抽出された各値についての長期バイアス成分の削除と正規化を行う。この正規化を行った値はアバターのリアルタイム動作制御情報として用いられる(4.3 節)と共に、非言語意図の表出イベントを判別する第二段階処理の入力として用いられる。

顔情報を正確な 3D モーションに用いようとする、カメラ位置と顔位置の関係のキャリブレーションを行わねばならないが、今回の目的では短時間で行われるアクションを取得することが主眼である。そのため、数十秒単位で変化のない定常的な成分は指数移動平均処理で抽出し、この基本バイアス成分からの差分となる数秒単位の挙動のみを意図的行動に相当するデータ成分として抽出する。これによりカメラとユーザの位置関係など初期条件に依存する要因や、カメラ前で体をうごかしたために顔位置のデフォルト場所が変化するなどの長期的な変化に起因する要素を削除している。

この 1 次指標では、頭部の回転運動(yaw / pitch / roll)、頭部の左右移動や前後移動、眉の上げ下げ、口の開閉などが抽出され、アバター挙動のリアルタイム制御情報として送られる。

4.2.2 2 次指標 (意図表出イベントの判別)

第二段階処理では、第一段階処理でフィルタ処理された値を閾値で分割して状態を定義し、またその遷移モデルとしての複合条件を用いてジェスチャや意図表出の有無の判別に用いる。このイベントが発生したと判断された場合には、特定の意図メッセージとしてシステム上の表出処理とラベル付けのロギングを行う。

ジェスチャアクション KPI :

顔位置情報から検出するジェスチャアクションには以下の三種類例を用いる。

アクション	判定手段
賛同(agree) 頷き	縦方向(pitch)の運動の遷移回数
拒絶(deny) 首振り	横方向(yaw)の運動の遷移回数
困惑(wondering) 首傾げ	顔方向軸周りの回転運動(roll)の角度

ジェスチャを定義する表現形態として、(1)状態を遷移した回数で定義するものと、(2)ある特定の状態にとどまる時間で定義されるパターンがある。頷きジェスチャや首振りのジェスチャは、所定時間内に発生する状態遷移の回数で定義され、困惑ジェスチャは特定の状態にとどまる時間で定義されている。ある条件を満たすとメッセージ発生として検出し、離散イベントとしてシステム内での通知を行う。このイベントは顔モデルのパターン変化や反応意見のカウント作業に用いるもので、いったんイベントが発生すると、既定の時間が経過するまで同一イベントは発生しない。

(これら判別に用いる時間と遷移回数のパラメータについては、ジェスチャ判別に対する鋭敏性にかかわる問題であるため、個人パラメータとして調整が必要になる。)

この意図表出段階では、ジェスチャから検出されたものと、表情から検出されたものがそれぞれ利用される。

顔表情 KPI :

本システムで利用する faceapi.js 等では Ekman 提唱の基本感情分類[4]に基づいて happy / sad / surprised / angry / disgusted/ fearful / neutral の8種の感情分類が行える。だが、このうち、angry / disgusted / fearful 等の表情は微表情として発生することが多く、特に日本人はこのような大きな表情表現を行うことが少ない[13]。わざと大きな表情を出した場合の検出は行えるが、そのような大きなネガティブ感情の表情は会話中でメッセージとして意図的に発生させることも少ないため、本システムではアバター操作の対象とせず、笑顔(happy)/ 悲しみ(sad)/ 驚き(surprised) の三種のみを反応アクションとして利用している。

4.3 アバター表示部モジュール

4.3.1 アバター表示モデル

アバター表示モジュールは、以下の二種類の表示モデルの表示に対応している。

個人用 3D モデル：登録ユーザ本人の顔画像を用いて作成した GLB 形式の 3D 胸像型モデルを表示する (図 5)。図の事例は Ready Player Me [16] で生成したモデルの取り込み事例である a。



図 5 3D モデル表示 (Ready Player Me での生成モデル)

Fig. 5. 3D avatar models from Ready Player Me

a Ready Player Me は、ユーザの顔写真から様々なメタバースアプリケーションで使える 3D アバターを作成する無料サービスであり、一般的なメタバースで広く使うことができる GLB モデルを提供する。

匿名ユーザ用画像モデル：2次元画像をベースにした表示モデル。顔部分、目、口など各画像をビルボードとして表示し、関節周りの回転やサイズを調整することで表情変化を行う(図 6)。



図 6 匿名ユーザ用画像モデル表示と漫符表現

Fig. 6. Paper avatar for anonymous user

4.3.2 吹き出しと漫符表現

今回のシステムでは、画面上の主役は共有ドキュメントでありアバターは画面内に小さく表示されることを前提としている。このため、アニメーションで細かい表情が出たとしてもそれが伝わるとは限らない。このため、笑顔や頷きジェスチャなどイベントとして取得された動作については、アバター上に漫符表現で表示する(図 6)。また、匿名アバターでは、賛成や笑顔のポジティブ反応は暖色で、当惑や否定などのネガティブ反応は寒色にするなどアバターの色で表現している。

5. コンテンツの共同視聴

遠隔会議では、スライドや文書などのコンテンツを共有して、閲覧箇所や指示箇所をリアルタイムで伝えあい、対話の補助に使う。

本システムではこのコンテンツ共有機能をモジュールとして用意し、意図表出イベントの発生とコンテンツの特定箇所を時刻同期してロギングすることで分析用データを生成する。

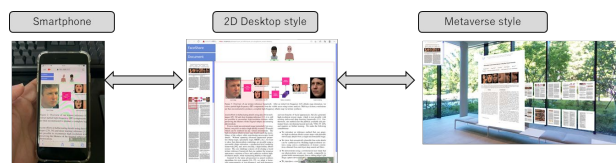


図 7 閲覧モードと視聴個所の共有

Fig.7. Asymmetric view-mode and viewpoint sharing

ディスプレイ領域の多様性に対処するため、2D 画面形態でコンテンツのみを表示するモードと、3D 画面表示形態で表示するモードの二種類がある。(図 7)。本章では、このような 3D シーン形式の表示と 2D ドキュメント形式の表示を自由に遷移して表示できる環境において、前章の顔画像とアバターによる反応情報を適切に扱う方法論を論ずる。

5.1 共同視聴の対象メディア

コンテンツ表示部は、PDF などの一般メディアを Web 上に表示して閲覧する機構で、2D 形態での描画と 3D 形態での描画機能と閲覧箇所の共有機能とログイン機能をもつ。共同視聴行為に用いるコンテンツのメディア例には以下のようなものがある。

- デスクトップ共有：WebRTC 機能を用いて、デスクトップ画面をリアルタイムに映像として送信し画面の共有を行う。この WebRTC 連携部の試作にはシグナリングサーバーが必要となるため Skyway (NTT) [17] のサービスを用いて実装を行なった。
- 静的書類の共有：PDF 文書の共同閲覧を行う。デスクトップで話者がページを特定しながら見ることもあるが、視聴者が部分を拡大したり別ページを閲覧視聴したりすることも多い。この PDF のレンダリング処理には OSS の pdf.js[18]を使用した。

5.2 顔表出と共同視聴の対応付け

5.2.1 議論コンテキストと視聴状態モード

顔画像から得られる非言語メッセージを遠隔コミュニケーションに取り込む場合、議論のコンテキストと一致して分析・表出を行えていることが前提となる。だが、視聴者が浮かべている表情が常に現在の議論コンテキストに向けたものとは限らない。

発言者は全員がデスクトップ共有された同じ画面を視聴している気になるが、実際には視聴者がコンテンツの同じ場所を視聴しているとは限らない。例えば話を聞きながらも別のページ箇所を確認しているなどの状況はしばしば発生するし、特に小さなスマートフォンなどを用いている場合には、各自がスワイプ拡大や別の手段でドキュメントを見ているなどの事象も自然と多くなる。

これが実世界であれば、ある表情を浮かべている聞き手が、説明者を見て話自体に集中して聞いているのか、手元のスライドを読んでいるのか、あるいは資料を参照しているのかといった動作からその表情の文脈を読み取ることができる。だが、リモート環境でカメラから機械的に表情反応を抽出すると、それが説明者に向けた表情なのか、資料の別ページをチェックしている際の表情なのかといった背景文脈が失われ誤解が発生する可能性がある。

特に、デスクトップ表示を前提としたスライド視聴に対し、PDF は印刷を前提とした資料が多く、読もうとすると拡大表示する必要がある場合が多く、文書の別箇所を参照することも多い。このような状況を適切に管理するために、PDF 共有機構では、視聴箇所を共有管理を行っている。

本提案では、共同視聴に当たってのユーザ行動を以下のようにいくつかのモードに分類して考える。

探索 (Walkthrough) :

ポスターセッションのように複数の会話が行われている空間において、概要を確認している状態。興味を持つかどうか、話し手の会話内容を聞くと同時に、聞き手の反応も判断の要素とすることがある。

会話合流 (Talk share) :

特定のグループに加わって話を聞いている状態。

たまたま立ちこのような視聴者は必ずしも時間軸にそった議論の文脈に合流できているとは限らない。

共同視聴 (page share) :

資料のうち話し手が説明するページを同期視聴し、話し手の発話を聞いている状態。話し手がポインタを用いてページの説明を行っている場合、聞き手はポインタの指す位置を追跡して発話内容の理解に務める。

部分限定視聴 (focus) : 提示されたページの理解、特に文章の一部について読解を行っている状態。

特にスマートフォンなどの小さい画面で視聴している場合には、文字読解のために画面を拡大して見るケースが多く、どの部分を読んでいるのかというアテンション情報の理解に使うことができる。

逸脱視聴 (looking aside) : 現在リアルタイムで行われている会話の時間軸から逸脱し、予稿などの文章を参照して前後の別の内容についての閲覧を行っている状態。

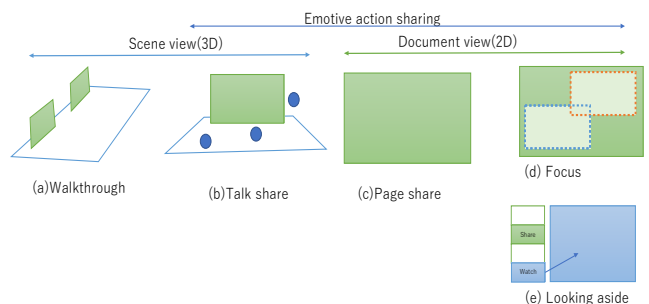


図 8 視聴形態の遷移と共同視聴の分類

Fig. 8. Transition of viewer mode and content sharing.

5.2.2 閲覧箇所の共有とログイン

提案システムでは、コンテンツモジュールの情報から視聴者がどの状態モードにいるかを特定し、視聴者の意図表出イベントの反応情報と併せて用いることによって、話し手への適切な非言語フィードバックを行うとともに、評判分析といった目的への応用を視野に入れてデザインを検討する。システムでは視聴ユーザの表示モードに合わせて、対応アバターの表現(サイズ・透明度)顕著度表現を調整する。またログデータの解釈に重みづけとして用いる。

視聴対象のコンテンツは document_id とそのサブ分割区分 (page_id) および、視聴領域によって中止内容が定義され、ログインが行われる。会話合流(Walkthrough)が行われた段階で document_id が定義され、2D 表示でドキュメント

が共有されたら page_id が使用される。また、別ページを参照した場合や、特定部分を拡大視聴した場合には、それぞれ部分限定視聴、逸脱視聴として取り扱う。

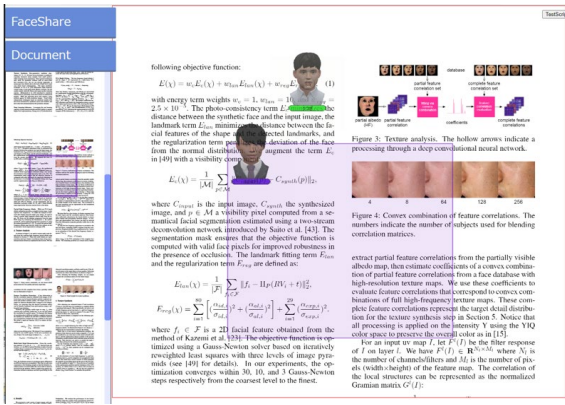


図 9 PDF の共同視聴画面の例

Fig.9. PDF sharing view

図 9 で示す事例では、紫の領域は顔アバターで示されるユーザが部分限定視聴を行って拡大視聴箇所を拡大している「部分限定視聴」の様子をリアルタイム共有するものである。また、別ページを見る逸脱視聴となった場合には、顔アバターの表示が行われなくなり、該当するページにアテンション情報が表示されるなど、話者に聴衆側の注目状況をリアルタイムで伝えている。

5.3 感情イベントの投票

2次指標で、ジェスチャや表情のイベントが発生した場合には投票処理が行われる。図 10 に、デモ内での感情イベントのカウンティング事例を示す。



図 10 感情イベントのカウンティング事例の例

Fig.10 Counting of emotional event in demonstration

6. 検証

本システムの実用化に向けての構成と効果に関する検証は現在計画中である。本節では要素モジュールについて幾つかの検証実験について説明する。

6.1.1 認識精度検証

システムの単体性能評価として、ジェスチャと表情によるイベント認識精度の計測を行なった。

図 16 にテスト時の提示情報例を示す。画面上に特定のジェスチャまたは表情を指示した期間(task)と反応を行わない期間(rest)を提示し、その認識を、アバターの表示フィードバックがある場合とない場合のそれぞれで計測した。

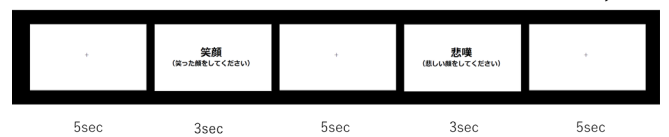


図 11 実験パターン例

Fig.11. Example of the experiment scheme

ジェスチャ認識

図 12 にジェスチャ動作の取得データのサンプルを挙げる。左は 1 次フィルタ処理を行なったジェスチャデータの例 (nod=上下運動, yaw=横運動, tilt=傾げ運動) 右は 2 次フィルタ処理を行なった後の KPI の例となる。

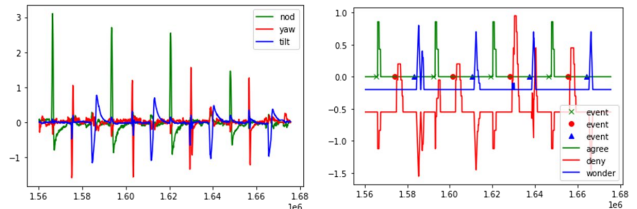


図 12 ジェスチャ認識のサンプル例

Fig.12. Sample of gesture recognition

質問に対して頷く、首を振るなどのジェスチャを行うように指示を行った。簡易実験ではジェスチャの検出精度は 91% となっている。但しこの検出は自然状態の頷きを取得するのではなく、あくまで UI としてのジェスチャを意識的に行ったものであることに留意する必要がある。また開発環境での小規模実験であるため、Web カメラを用いてノイズの多い環境の場合の検証には、さらなる大規模な実験データが必要とする。

感情認識

図 13 に 6 種類の感情を指示して行なった場合の KPI (一次フィルタ済み処理情報) の例を示す。特に「怒り(angry)」「嫌悪(disgusted)」「恐怖(fearful)」などは、被験者も指示に対応する表情を安定して作ることができず、またそのため他の感情との区別が非常に難しく、相互の誤検出がしばしば発

生ずる（誤検出をグラフ内の吹き出しとして表記）。これらの条件から、システムの操作には表情実施がしやすく検出にも安定性のある、笑顔(happy)/悲嘆(sad)/驚き(surprised)の三種類のみをアバター操作対象のイベントとして利用している(図 14)。

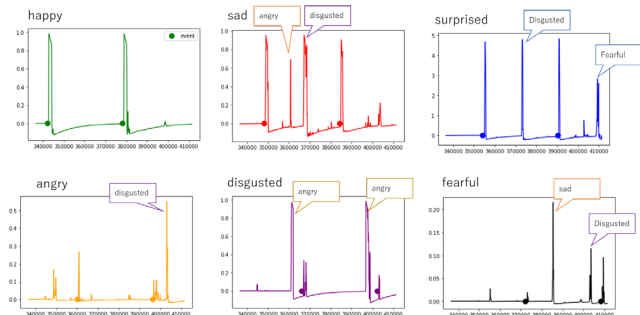


図 13 6種の感情表現を用いた場合の誤検出事例

Fig 13. Error pattern of 6 emotion model

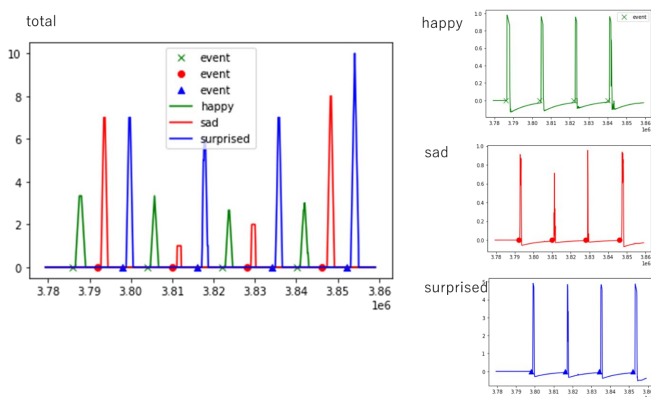


図 14 3種の感情表現での分離事例

Fig 14. Emotion recognition (3 pattern)

7. 結言

本発表では遠隔コミュニケーションに対してメタバース型の参画をするためのシステムの方法論についての議論を行った。

本試作では、会話に対して顔表情による操作に基づく非言語情報のフィードバックを行う機構の実現と、同システムに多様な接続機器で接続した場合のユーザビリティを担保するための画面構成デザインについて議論を行った。

異種機器の環境においてディスプレイの面積に基づく行動差異が発生することを前提として、共有ドキュメントの視聴行動を複数の参照モードに分けて管理する。視聴共有と部分注目、逸脱参照といった状態を明示化管理することにより、表情の変化からメッセージを発信する機構の誤解が生じないように設計されている。

今後の検討事項として、実運用のユースケースと実例デ

ータの収集と併せて、異種環境による画面構成を用いたことによる効果検証を検討中である。

参考文献

- [1] Smart, Cascio, and Paffendorf, "Pathways to the 3D Web: A Cross-Industry Public Foresight Project." Metaverse Roadmap. <https://www.metaverseroadmap.org/MetaverseRoadmapOverview.pdf> (2007)
- [2] Meta, "Connect 2021 : メタバースに向けたビジョン" (2021. 10)
- [3] Microsoft Japan News Center (2021. 11), "Mesh for Microsoft Teams が目指すメタバース空間でのより楽しく、よりパーソナルなコラボレーション", <https://news.microsoft.com/ja-jp/2021/11/04/211104-mesh-for-microsoft-teams/>
- [4] Ekman, P., Friesen, W. V., "Unmasking the Face: A guide to recognizing emotions from facial expressions", Malor Books (1987)
- [5] Sasaki, W., Nishiyama, Y., Okoshi, T. et al. Investigating the occurrence of selfie-based emotional contagion over social network. Soc. Netw. Anal. Min. 11, 8 (2021).
- [6] Coles, N. A., Larsen, J. T., & Lench, H. C. (2019). A meta-analysis of the facial feedback literature: Effects of facial feedback on emotional experience are small and variable. Psychological Bulletin, 145(6), 610–651.
- [7] 串田秀也, "聴き手による語りの進行促進 — 継続支持・継続催促・継続試行 —", 認知科学, 16(1), pp. 12-23.(2009).
- [8] Bavelas, et al. Listener as co-narrators. Unpublished manuscript, Department of Psychology, University of Victoria (2000).
- [9] 折重将人, Samratul Fuady, 李浩夔, 三武裕玄, 長谷川晶一, "アバターロボットによる存在感通信について", VR 学会大会論文集 (2016).
- [10] 西村優, 佐中村泰, 石黒浩, "敵対的生成ネットワークを用いた対話中の人間の振る舞いのモデル化", 日本ロボット学会誌 Vol. 37. No7. pp632-638, (2019).
- [11] 吉田有花, 宮下芳明: 身体動作の重量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).
- [12] 松長雄也, 谷中俊介, 坂内祐一, "動画視聴における他視聴者の音声情報の再生方式", DICO 2018, pp530-533, (2018)
- [13] Wataru Sato, Sylwia Hyniewska, Kazusa Minemoto and Sakiko Yoshikawa. Facial Expressions of Basic Emotions in Japanese Laypeople. Frontiers in Psychology, 10:259. (2019)
- [14] Cozzi, Patrick (2015). "WebGL Engine Design in Babylon.js". WebGL Insights. CRC Press. ISBN 978-1498716079.
- [15] Vincent Mühler, <https://github.com/justadudewhohacks/face-api.js/>
- [16] "Ready Player Me: Metaverse Full-Body Online 3D Avatar", <https://readyplayer.me/>
- [17] "Skyway - アプリや Web サービスに、ビデオ・音声通話をかんとんに導入・実装できる SDK", <https://webrtc.ecl.ntt.com/>
- [18] "PDF.js - A general-purpose, web standards-based platform for parsing and rendering PDFs", <https://mozilla.github.io/pdf.js/>
- [19] 宇都木契, 藤原貴之, 小林美保, "視聴者反応をフィードバックする Web VR 環境を用いた遠隔ポスター発表システム", DICO 2021, pp. 1636-1639