対話者顔方向検出に基づく自己キャラクタ対面合成による身体的ビデオチャットシステムの開発

高田 友寛[1,a)]　中山 志穂[1]　石井 裕[2,b)]　渡辺 富夫[2,c)]

概要：著者らはこれまでに，ビデオ映像を用いた遠隔コミュニケーションにおいてインタラクション把握を支援する手法として，自己の代役キャラクタを相手映像に対面合成する実映像対話システム EnhancedVideoChat(E-VChat) を開発してきた．E-VChat では対話相手が自己を正面から撮影することを想定してキャラクタを配置していたが，実際の利用場面では横方向などから撮影する場面も考えられる．本研究では，E-VChat において対話者の顔を検出し顔方向を推定することで，相手が不特定の箇所にカメラを設置している場合においても画像処理により相手の対面に自己キャラクタを重畳合成する実映像対話システムを開発している．

Development of a Video Chat System in which Talker's own Character is Superimposed for a Face-to-Face Projection based on the Detection of the Partner's Face Angle

Tomohiro Takada[1,a)]　Shiho Nakayama[1]　Yutaka Ishii[2,b)]　Tomio Watanabe[2,c)]

Abstract: We have proposed an embodied video chat system in which own avatar is superimposed on the other talker's video images, in order to comprehend talker's mutual interaction in remote communication, called E-VChat system. Our previous system was developed from the only frontal viewpoint, though talkers place the camera on an indefinite position in practical use. In this paper, we develop an embodied video chat system in which a talker's avatar is superimposed on a face-to-face projection based on the detection of the partner's face angle from any viewpoint of camera by image processing.

1. はじめに

今日，ビデオチャットなどの実映像を用いた遠隔コミュニケーションの利用が盛んになっている．従来よりビデオチャットでは相手映像のみか，あるいは自己映像を P-in-P で挿入する手法などが用いられているが，これらの手法では画面構成が分離されており互いのインタラクションが把握しにくいといった問題がある．一方で，人は対面コミュニケーションにおいて言葉によるバーバル情報だけではなくうなずきや身振り・手振りなどのノンバーバル情報を用いて互いに引き込み合うことで円滑なコミュニケーションを行っており，身体性や身体的引き込みなどのリズム同調に着目した様々なコミュニケーション支援手法が検討されている [1]–[4]．特に，著者らはこれまでにインタラクション把握を支援するビデオコミュニケーション手法として，自己の代役となって引き込み動作を行うキャラクタを相手映像に対面合成する実映像対話システム EnhancedVideoChat(E-VChat) を開発してきた [5]．E-VChat システムでは Web カメラがモニタ上部に設置されている場合を想定して自己の代役となるキャラクタを画面下方中央に配置していたが，実際の利用場面ではカメラ

[1] 岡山県立大学大学院 情報系工学研究科
Graduate School of Systems Engineering, Okayama Prefectural University, 111 Kuboki, Soja-shi, Okayama-ken 719-1197, Japan
[2] 岡山県立大学 情報工学部
Faculty of Computer Science and Systems Engineering, Okayama Prefectural University
a) t_takada@hint.cse.oka-pu.ac.jp
b) ishii@cse.oka-pu.ac.jp
c) watanabe@cse.oka-pu.ac.jp

をモニタ横に設置したり，自己を横方向から撮影する場合も想定される．そのような場合でも相手正面で自己のキャラクタを対話相手にできるだけ正対して配置することで，互いのインタラクションを把握しながらの遠隔対話を行うことができると考えられる．また E-VChat システムでは，これまで頭部動作計測デバイスを頭部に装着することで対話者の頭部動作を計測し，自己キャラクタに対話者の頭部動作を連動させキャラクタが自己の代役であることを明確にしていたが，装着感の煩わしさやデバイスがずれるなどの問題があった．

本研究では，E-VChat システムにおいて画像処理によるフェイストラッキングを用いることで，対話相手の視線の先に自己キャラクタを配置し場面に応じてインタラクション把握を支援する実映像対話システムを開発している．本システムは，検出した顔の角度情報を用いて自己の頭部動作を画面上の自己キャラクタに反映させることで，デバイスを装着することなくキャラクタに対話者の頭部動作を連動させることができる．

2. E-VChat システム

2.1 システム概要

ビデオコミュニケーションにおいて互いのインタラクション把握を支援する手法として，著者らはこれまでに EnhancedVideoChat(E-VChat) システムを開発している．E-VChat システムでは，自己の代役となるキャラクタを画面上の相手と対面するように重畳合成し仮想的に対面コミュニケーション場を形成することで，相手と自己とのインタラクション把握を容易にすることができる (図 1)．自己キャラクタには音声リズムに基いてうなずきや身振り手振りなどの話し手・聞き手動作を自動で行わせ，身体的リズムの共有を助けるとともに自己に対して共感反応をフィードバックし，話者の会話意欲を促進する．自己キャラクタには音声に基づく自動動作に加え，ヘッドセットタイプのモーションキャプチャデバイスを用いることで話者の頭部動作を連動し，肯定・否定といった意思表示を反映することでキャラクタが自己のアバタであることを明確にしている [6]．また，自由対話を想定した官能評価実験によりシステムの有効性が示されている [5]．従来の E-VChat システムではキャラクタに対話者の頭部動作を連動させるため，頭部動作計測デバイスを対話者の頭部に装着する必要があったが，この方法は装着感などで対話者に違和感を与える可能性がある．そのため，本研究では Kinect (Microsoft©) [7] を用いて画像処理による頭部動作検出によるシステムを開発した．対話者にデバイスを装着する必要性をなくすことで，この装着感の問題を解消している．

2.2 音声による自動生成に基づくのキャラクタ動作モデル

自己の代役としてのキャラクタには，入力された音声の ON-OFF パターンからコミュニケーション動作のタイミングを推定し，聞き手動作，話し手動作などの身体動作を行わせる．音声データは 16bit 22.050kHz でサンプリングし，閾値で二値化するとともに，音節間の短時間の無音区間による発話の断片化を除去するために 133msec でハングオーバ処理を施している．

図 1 E-VChat システム

聞き手動作においては，音声の ON-OFF パターンに基づくうなずき反応モデルと，腕部および上体部に対してうなずきの予測値に基づく身体動作モデルを用いている [8]．うなずき反応の予測モデルはマクロ層とミクロ層からなる階層モデルで構成している．マクロ層では音声の呼気段落区分での ON-OFF 区間からなるユニット区間にうなずきの開始が存在するかを [i-1] ユニット以前のユニット時間率 R(i) （ユニット区間での ON 区間の占める割合，(2) 式）の線形結合で表される (1) 式の MA (Moving-Average) モデルを用いて予測する．予測値 Mu(i) がある閾値を越えて，うなずきが存在すると予測された場合には処理はミクロ層に移り，音声の ON-OFF データ（30Hz，60 個）を入力とし，(3) 式を用いて MA モデルでうなずきの開始時点を推定する．

また，話し手の身体動作においては，音声の ON-OFF パターンに基づいた身体動作モデルの MA モデルを用い，頭部動作よりも低い閾値に基づいて，腕や胴体などの各部位を動作させる．

$$M_u(i) = \sum_{j=1}^{J} a(j)R(i-j) + u(i) \quad (1)$$

$$R(i) = \frac{T(i)}{T(i)+S(i)} \quad (2)$$

$T(i)$: i 番目のユニットでの ON 区間

$S(i)$: i 番目のユニットでの OFF 区間

$u(i)$: ノイズ

$$M(i) = \sum_{j=1}^{K} b(j)V(i-j) + w(i) \quad (3)$$

$b(j)$: 予測係数

$V(i)$: 音声データ

$w(i)$: ノイズ

2.3 キャラクタ重畳合成の手法

実映像コミュニケーションにおいて，図 2 に示すように

カメラをモニタ上部に設置する場合は，カメラは自然とモニタを見る対話者を見下ろす角度となる．このときモニタを注視している話者と画面上の対話相手との間には視線のずれが生じるが，画面の下部中央にキャラクタを配置することで，正面から撮影される対話相手が自己キャラクタと対面しているように見える (図 3)．これにより，視線のずれによって生まれる画面上の違和感を解消することができ，対話相手と自己の代役キャラクタとの関係性を視覚的に把握することで，互いのインタラクションを把握しながらの円滑なコミュニケーションを行うことができる．

カメラをモニタ上部に設置した場合の E-VChat システムにおいて，キャラクタを重畳合成する手法の有効性が示されていることから [5]，カメラをその他の場所へ設置した場合においても自己の代役キャラクタを対話相手の視線の先へ配置することで同様の効果が得られることが期待される．

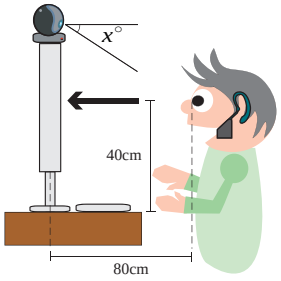

図 2　視線位置模式図

図 3　画面上での目線

3. 視線の先へキャラクタ移動を行う E-VChat システム

3.1 コンセプト

著者らがこれまでに開発してきた E-VChat システムでは，Web カメラをモニタ上部に固定する構成を前提としてきた．しかし，現実的な利用場面を考慮すると，対話相手がカメラをモニタ上部に設置しているとは限らないため，カメラをモニタ横に設置していたり横方向から撮影する場合についても想定する必要がある．

そこで本研究では，Kinect を用いて対話者の顔の位置と角度を検出し，顔向きの先へ自己キャラクタを移動させることで，対話相手が Web カメラを不特定の場所に設置している場合においても自己キャラクタを介して円滑なコミュニケーションを行うことができる実映像対話システムを提案する．

システムのコンセプトを図 4 に示す．対話相手を Kinect を用いて撮影した相手映像と，自己の代役として自動動作を行うキャラクタを PC の画面上で重畳合成する．相手映像と自己キャラクタを仮想的に対面合成することで，対話者が互いの身体的インタラクションを捉えやすい状態を実現し，ビデオ映像を用いて対話相手の表情や身体動作などのノンバーバル情報を観察しながら会話を行うことができる．キャラクタは対話者の発話音声に基づいて「話し手」としての動作を行い，対話相手の発話音声に基づいて「聞き手」としてうなずきや身振り手振りなどの動作を行う．

また，Kinect を用いて対話者の顔の位置や向きを計測して対話相手の視線の先を推定し，推定位置に自己キャラクタを移動させ，画面上の対話相手と対面するように配置する．これにより，対話相手が正面を向いていない場合においてもキャラクタを介して自己と相手とのインタラクションを把握しやすい状況をつくることができる．

さらに，対話者の顔計測データを用いて，自己キャラクタに対話者自身の頭部動作を連動させる．自己キャラクタに頭部動作を連動させることで，キャラクタが自身の代役であることを明確にするとともに，肯定や否定といった意志表現を反映可能にすることで，よりキャラクタを介したコミュニケーション特性を生かしたシステム構築を行うことができる．

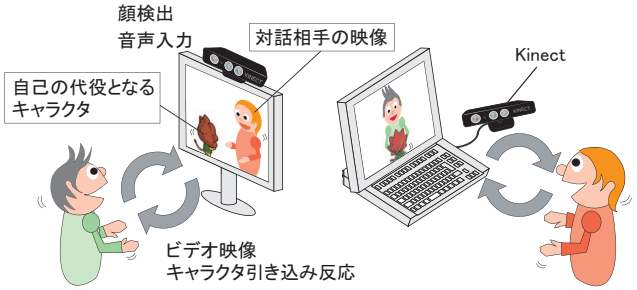

図 4　コンセプト

3.2 システム構成

開発したシステムの使用風景を図 5 に示す．システム構成における対話者間は 1GB/s のイーサネットで接続されており，音声通信を行う．顔情報の検出には Kinect for Windows センサ（L6M-00005) を用いている．図の使用場面では，本システムを用いたコミュニケーションの一例として，対話者・対話相手両者ともにモニタ横に Kinect を設置し，顔を斜め下から撮るという構成をとっている．モニタ画面上には対話相手の映像に自己の代役となるキャラクタを重畳合成しており，対話相手の顔の角度に応じてキャラクタを移動し，対話相手の方向を向くようにキャラクタを回転させる．また，画面に対する顔の位置にも対応してキャラクタを移動させることで，自己キャラクタと対話相手が対面するようにキャラクタを自動で適切な配置へ移動させることができる．これにより，画面上の対話相手が不特定の方向を向いていても，顔を認識できる範囲であればキャラクタを対話相手と対面するように移動させることが可能となっている．

3.3 顔方向推定手法

画面上の対話相手の顔方向については，Kinect のフェイストラッキングを用いて取得した顔の 6 軸情報（3 軸角度・

図 5 システムの使用風景

3 軸位置情報）を用いて推定する．位置情報の前後方向については赤外線センサにより計測した深度情報を取得している．対話相手がカメラに対して正面を向き，自己キャラクタが画面中央下方に位置して奥方向を向いてる構成を初期位置とする．顔の 3 軸角度情報の内，Yaw 角に対応して自己キャラクタを回転させ，同時に x 軸移動を行う．また，顔の位置情報にも対応し，対話相手と対面するのに適切な位置へキャラクタを移動させることができる．

本システムは対話相手がどのような場所に Kinect を設置していても自己キャラクタを適切に配置することを目的としているが，フェイストラッキングが可能な範囲には限界があるため，Kinect に対する対話相手の顔の角度には制限がある．図 6 に上下左右の顔角度検出における限界範囲を示す．フェイストラッキングでは両目・鼻・口などの特徴点をトラッキングするため，片目が隠れる角度などは顔を正常に検出することができなくなる．本システムでは，顔検出が正常に行えなくなった場合には，対話相手はおおよそその方向を向いていると仮定し自己キャラクタをその場で停止させる手法をとっている．

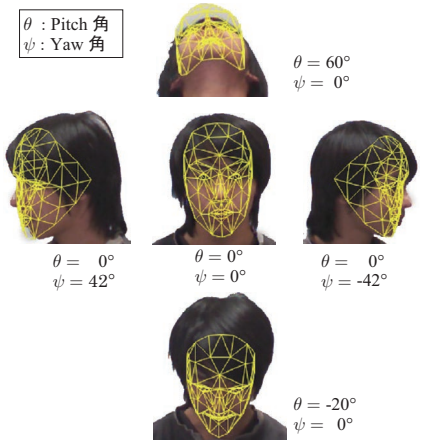

図 6 顔角度検出範囲

3.4 キャラクタ移動手法

相手視線の先へのキャラクタの移動手法については，(I) キャラクタを常に適切な場所へ移動させる手法，(II) 対話者がキー入力した際に移動させる手法，(III) 顔方向に一定時間同一方向を向いている場合に移動させる手法などが考えられる．(I) の手法ではキャラクタは忙しなく動き回ることになり，画面を見ている対話者に不自然な印象を与え対話を妨げる可能性が考えられる．(II) の手法では対話者がキャラクタを移動させたいときに移動させることができるが，対話者が意図的に自己キャラクタを操作すること自体がコミュニケーション支援の妨げになることが考えられる．そこで，本システムの初期設定では (III) の「顔方向に一定時間同一方向を向いている場合に移動させる手法」を選定している．対話相手が一定時間同一方向を向いている場合のみキャラクタを移動させることで，対話相手の首振りや遠くの物を見る動作などに左右されることなく，対話者が基準として向いている先へキャラクタを移動させることができる．また，本システムでは対話者が使用したいモードを選択できる構成となっており，用途によって対話者にとって最適なキャラクタ移動手法が異なる場合にはモードを変更することができる．

4. おわりに

本論文では，相手が不特定の箇所にカメラを設置している場合においても相手映像に自己の代役となるキャラクタを対面合成して相手の視線の先へ自動移動させる E-VChat システムを開発した．本システムと提案手法が実映像コミュニケーションに及ぼす影響等についての検討は今後の課題である．

参考文献

[1] 渡辺 富夫: 身体的コミュニケーション技術とその応用; システム/制御/情報, Vol.49, No.11, pp.431-436, (2005).

[2] 森川 治, 橋本 佐由理, 前迫 孝憲: 仮想的な抱擁を取り入れた遠隔カウンセリングシステム，日本バーチャルリアリティ学会論文誌，Vol.14, No.1, pp.3-10, (2009).

[3] Otsuka, K.: Multimodal Conversation Scene Analysis for Understanding People's Communicative Behaviors in Face-to-Face Meetings, Human Interface, Part II, HCII 2011, LNCS 6772, pp. 171–179, (2011).

[4] 石井 亮，宮島 俊光，藤田 欣也: アバタ音声チャットシステムにおける会話促進のための注視制御，ヒューマンインタフェース学会論文誌，vol.10, No.1, pp.87-94, (2008).

[5] Takada.T, Ishii.Y, Watanabe.T: Development of an Embodied Video Communication System with a Superimposed Entrainment Character Driven by Voice and Head Motion Inputs, Proc.of First International Symposium on Socially and Technically Symbiotic Systems,No.40,pp.1-4(2012).

[6] Yamamoto, M., Osaki, K., Matsune, S., and Watanabe, T.: An Embodied Entrainment Character Cell Phone by Speech and Head Motion Inputs, Proc. of the 19th IEEE International Symposium in Robot and Human Interactive Communication Symposium, pp.318-323 (2010).

[7] Kinect for windows SDK from Microsoft Research. http://kinectforwindows.org/.

[8] Watanabe, T., Okubo, M., Nakashige, M., and Danbara, R.: InterActor: Speech-Driven Embodied Interactive Actor; International Journal of Human-Computer Interaction, Vol.17, No.1, pp.43-60 (2004).