

発話がぶつからない Web 会議を実現するための発話欲求伝達手法

玉木 秀和^{1,a)} 東野 豪¹ 小林 稔¹ 井原 雅行¹

受付日 2012年4月24日, 採録日 2012年10月10日

概要: 本研究では, 低コストで利用場所の制約が少ない Web 会議システムを用いて司会者のいないアイデア出し会議やフリーディスカッションを実施する際に, 発話の衝突を低減させ, 円滑な話者交替を可能にすることを目的とする。そのために, 本稿では発話欲求伝達手法を提案する。まず参加者の動作をシステムが観察し, その参加者の発話欲求度合いを推定する。参加者各々に, その推定結果をリアルタイムにフィードバックすることで, 参加者自身が動作を調整し, システムによる推定結果と実際の発話欲求度合いが近づいていく。推定された発話欲求度合いが設定された閾値を超えた場合, その参加者の映像を強調して他の参加者へと知らせる。本手法を実現するプロトタイプを実装して評価実験を実施した結果, 提案手法により発話欲求を他の参加者へ伝達でき, 発話衝突を低減させることができることが分かった。

キーワード: 遠隔会議, 話者交替, 発話衝突

A Method Communicating Desire of Speaking to Avoid Speech Contention in Web Conferences

HIDEKAZU TAMAKI^{1,a)} SUGURU HIGASHINO¹ MINORU KOBAYASHI¹ MASAYUKI IHARA¹

Received: April 24, 2012, Accepted: October 10, 2012

Abstract: Our aim is to avoid speech contention in Web conferences. We propose a method communicating desires of speaking. Participants can check how strong their desires are detected with an indicator and adjust their motions to fit the indicator with their real desires. And if desire level increases over a threshold, frame of his/her picture is set color to yellow to indicate his/her desire to other participants. We took a brief test and investigated that our concept can avoid speech contention and realize a smooth turn taking in Web conferences.

Keywords: distributed conferences, turn-taking, speech contention

1. はじめに

CO₂ 排出量低減, 節電施策, 世界不況, パンデミック対策などが叫ばれ, 遠隔会議システムの需要が増加している。遠隔会議システムの中でも Web 会議システムは, デスクトップ PC やノート PC, タブレット端末で利用できるうえ, ソフトウェアのみで導入できるなどの手軽さが

あり, 市場も成長傾向にある [1]。本稿では Web 会議を, Web ブラウザ上で映像, 音声を共有しながら実施する遠隔会議と定義する。特に最近では災害発生時に移動が困難な場合に備え, 遠隔地間でのコミュニケーション手段の整備が急務であるとされる。遠隔地間で対策を練るための会議を実施する, もしくは企業活動の継続のために会議を実施する手段として Web 会議は有効な手段であるといえる。本研究は, Web 会議システムを用いて多地点間で司会者を設けないアイデア出しの会議 (創造会議 [2]) やフリーディスカッションを快適に実施できる環境の実現を目指す。

Web 会議はデスクトップ PC やノート PC, タブレット

¹ 日本電信電話株式会社 NTT サイバーソリューション研究所
NIPPON TELEGRAPH AND TELEPHONE CORPORATION,
NTT Cyber Solutions Laboratories, Yokosuka,
Kanagawa 239-0847, Japan

^{a)} h.tamaki@ntt.com

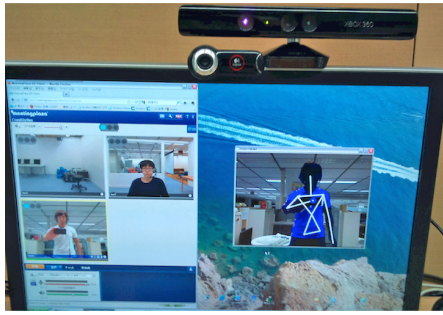


図 1 発話欲求を伝達可能な遠隔会議システム実行画面

Fig. 1 Display of the distributed conferencing system communicating the desire of speaking.

端末などを、多くの場合帯域の保証されないネットワークに接続して実施する。このため専用のインターネット回線や専用の部屋、大型のディスプレイを用いる TV 会議システムと比較し、個々の参加者の物理的な映像サイズが小さい、映像の解像度が低い、フレームレートが低い、リップシンクがとれない、映像、音声遅延が生じるなどの制約がある。これらに起因して他の参加者の様子を読み取りにくく [3]、誰がいつ発話し始めるのかを判断しにくい。そして、いざ発話を開始すると、他の参加者の発話と衝突してしまうことが多い。遠隔コミュニケーションにおいて音声遅延が 300 ms を超えると、特にこの発話の衝突が顕著になると報告されている [4]。また、Sacks らは発話衝突が生じた場合、その状態から復旧するために、同時に発話した参加者のうちいずれかが発話を中断すること、もう 1 度言い直すことなどが必要になると述べている [5]。このような状態では、素早い話者交替を頻繁に行う会議は実施しにくく [4]、参加者に負担をかける可能性がある。著者らの分析結果では、さかんに話者交替の起こる、司会者のいない創造会議を 4 地点で行う場合、Web 会議では対面会議と比較して発話の衝突が 30 倍近く起こることが分かった [6]。また、著者らは遠隔会議において発話衝突の与える影響を調査し、発話衝突が増加するほど精神的ストレスも増加することを明らかにした [7]。

人は普段、対面したコミュニケーションの場面では顔きや姿勢の変化など、非言語情報をうまく利用して発話の衝突を避け、円滑に話者交替している [8], [9]。しかし、Web 会議では先に述べた制約のため非言語情報を利用しにくく、発話の衝突が起こりやすい。

本研究の目的は、Web 会議において発話の衝突を低減させ、素早く頻繁に話者交替する会議を実施可能にすることである。著者らはこれまで、Web 会議において、人が発話の前に行う特徴的な動作（以降「予備動作」と呼ぶ）を検知して、最も次に発話しそうな参加者を選定し、全参加者へ示すことで話者交替を円滑化する手法を提案してきた [10]。システムが参加者の発話欲求を正確に把握して次の話者を示すことで、参加者の行動が自然に方向付けられ、

円滑な話者交替を行えることは理想的であるが、現在の技術ではそれは困難である。そこで本稿では発話を転換し、システムが参加者の発話欲求を把握する過程を参加者自身が手伝えることで、それを正確に把握させる仕組みを導入する。具体的には、システムによって推定されている発話欲求の度合いを参加者自身がインジケータでリアルタイムに確認し、それを基に自身の動作を調整し、インジケータのゲージがある閾値を超えたときに、明確なシグナルとして全参加者へと通知する手法（図 1）を提案する。

2. ビデオ会議と話者交替

映像コミュニケーションシステムの研究開発は古くから行われてきた。鑑沢らは、離れた 2 つの地点を結んで遠隔コミュニケーションにおいて、遅延のある環境では発話の衝突が多く起こることを報告した [4]。Sellen は多地点を結び、参加者映像を升目上に配置したデザインの遠隔コミュニケーションシステムを使用した際に、円滑な会話ができなくなることを指摘した [11]。その後通信技術や映像技術が急速に発達し、現代では高い解像度の映像を低遅延で送受信できる遠隔コミュニケーションシステムが実現してきている [12]。また、遠隔地にいる参加者の数だけのディスプレイを用意し、誰が誰を見ているかといった情報までを再現しようとする試みも行われている [13]。これらのように、遠隔地にいる参加者が、あたかも同じ部屋にいてテーブルを囲んでいるかのような状態を再現しようとするアプローチは非常にさかんである。しかし、高いコストがかかること、システムが物理的に大きくなることから、本研究で利用シーンとして想定しているような、多くの人々が気軽に、利用場所を選ばずに始められるような会議には適さない。

一方低コストで、どこからでも使える Web 会議システムも、映像や音声の質が向上してきている。ところが実際に帯域の保証されていないネットワーク上で Web 会議を行う際には、音声と映像両方の質を高く保つことは難しく、それらはトレードオフの関係になる。そのため音声の質を優先させた場合には映像のフレームレートや解像度を下げなくてはならない。その場合、参加者間で互いの非言語情報を読み取ることは非常に困難になるが、その問題を解消するためにさまざまな試行錯誤が行われている [14], [15], [16]。商用システムの MeetingPlaza [14] は話者交替をサポートするために、ボタンによりマイクのオン/オフを切り替える機能や、挙手アイコンを表示させる機能を実装している。石井らは参加者のアバタを CG で用意し、次の発話を促す視線を疑似的に表現する手法を提案した [15]。この手法では、音声を基に視線を向ける動作を疑似的に作り出している。Kawashima らは 2 者間の遠隔コミュニケーションにおいて、フィラ（間を埋める音声）を映像に置き換えて表現した [16]。これらは人とエージェントシステムとの対話

を円滑にすることを目的とする技術 [17], [18] を応用しているが、会話をする相手の実際の挙動を反映して伝えることはできない。

最近では、映像の解像度やフレームレートを、人が違和感を感じない程度に下げること、非言語情報が伝わる程度の 3 次元映像を送受信する興味深い試みも実施されている [19]。しかし、Web 会議において話者交替に必要な非言語情報を伝達でき、円滑に話者交替できる環境を実現した例はまだ存在しない。

3. 発話衝突を低減させるための基本アイデア

我々の研究目標は、Web 会議が持つ手軽さや、使用場所の制約の少なさは損なわずに、発話衝突せず円滑に話者交替し、気軽に質の高い遠隔会議ができるという、新しいコミュニケーションメディアを実現することである。より具体的には、人が話者交替をするときに使用する予備動作の伝達を支援し、円滑な話者交替を可能にすることを目指す。

図 2 のような話者交替時の認知・行動モデルを考えた。ある参加者の脳内で発話欲求が生じたとき、その参加者は外部へと何らかの予備動作を表出する。そして他の参加者の予備動作を観察し、自身が発話するのか、それとも沈黙して他の参加者へ発話を譲るのかを判断する。そして実際に発話、もしくは沈黙するといった行動に移る。対面コミュニケーションにおいて、このモデルの各ステップは滞りなく流れることが考えられる。しかし Web 会議においては、画像のサイズや解像度、ネットワーク遅延などの原因により、予備動作が他の参加者へ伝わりにくく、それを観察することが困難になる、と著者らは考える。

そこで図 2 の予備動作ステップと観察ステップの間に、何らかの仕組みを挟むことで、予備動作を他の参加者へと伝えやすくする手法を考える。具体的には、参加者の予備動作をシステムが検知し、検知した動作に基づいて参加者の発話欲求を推定し、それを他の参加者へと伝える方法をとる。すなわち、図 2 の「*」の位置に、図 3 のモジュール

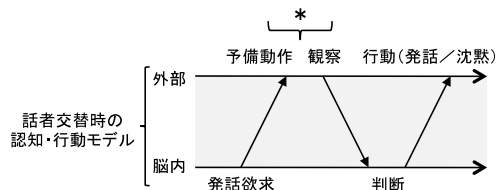


図 2 話者交替時の認知・行動モデル

Fig. 2 Cognition and action model in turn-taking.

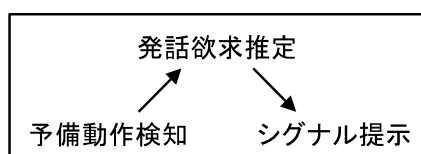


図 3 発話欲求伝達モジュールの処理

Fig. 3 Process in desire communicator module.

ルの処理を組み込むことを考える。図 3 のシグナルとは、参加者への注目を集めるための何らか（映像や音声など）の刺激を指す。

著者らはこれまで、Web カメラで取得した画像を基に、予備動作である可能性のある動作を検知し、そこから推定される、最も次に発話する可能性の高い参加者を、次話者候補として他の参加者へ強調して提示する手法を提案してきた [10]。現段階で、予備動作である可能性のある動作を検知した際に、それが発話欲求をとともなう予備動作であるか否かを正確に判断する方法は見出せていない。本稿では以降、予備動作である可能性のある動作を、予備動作候補と呼ぶ。

予備動作候補としてこれまで著者らがあげてきた、「傾き」「挙手」「手を顔周辺へ動かす動作」などをシステムが検知しても、前述のように、その動作をした参加者に発話欲求があるか否かを正確に把握することはできない。たとえば、「大丈夫」という意思を表す動作や、「賛成の人は？」などの質問に対して手を挙げる動作などは、必ずしもその後発話をともなうとは限らない。しかしこれらの動作と発話欲求のある「挙手」をシステムは正確に区別することはできない。したがって、ある参加者の発話欲求の度合いを正確に特定する手段は見出せていない。また、システムが次話者候補を 1 人に絞って提示する仕組みでは、発話欲求が正確に分からない場合、図 2 の判断ステップで、参加者に誤った判断をさせる可能性を多く含む。

4. 発話欲求伝達手法の提案

参加者の行う予備動作候補を正確に検知できたとしても、そこから発話欲求を正確に特定することは困難である。そこで著者らは発想を転換し、参加者へ推定されている自身の発話欲求度合いをリアルタイムにフィードバックする。参加者自身が動作を調整することで、自身のいだいている発話欲求とシステムに推定されている発話欲求の度合いをすり合わせることができ、発話欲求度合いと表示されるシグナルの整合率を高めることのできる、発話欲求伝達手法を提案する。本手法のポイントは、参加者が普段の会話で話者交替に使用している予備動作を検知して発話欲求の推定をすること、そしてシステムに推定された発話欲求の度合いと実際の発話欲求の度合いの整合性を、参加者自身が動きを調整することで近づけていく点である。

本手法を実現するために、以下の 3 つの項目を検討する。

- 1) 予備動作候補を検知する精度 (予備動作候補検知精度) の向上

発話欲求度合いを推定するために用いる、予備動作候補の検知精度を高める必要がある。そのために、センシングデバイスにより、参加者の頭、首、肩関節、肘関節などの 3 次元座標位置をセンシングし、その位置関係や動きを基に、それらの動作を検知する。センシングデバ

イスには、本稿では、2次元画像とその奥行き情報を高い精度で検出できる Kinect センサ [20] を採用する。

- 2) 検知された予備動作候補から推定される発話欲求を、実際に参加者のいづく発話欲求へと近づける仕組みの構築

その時々検知された予備動作候補を基に、システムに推定されている発話欲求度合いを表すインジケータ（発話欲求フィードバックインジケータ）を表示する。これにより、表示されている発話欲求フィードバックインジケータが、実際の発話欲求よりも高い場合、随意的に動作を抑え、発話欲求フィードバックインジケータを下げる事が可能となる。逆に、実際よりも発話欲求が低く推定されているとフィードバックされた場合には、発話欲求フィードバックインジケータを確認しながら予備動作を行うことで、推定される発話欲求度合いを引き上げることができる。人は普段の対面した会話において、発話欲求のある場合には何らかの予備動作をし、そうでない場合には動作を行わない。この予備動作はボタンを押す動作などとは異なり、普段対面会話において話者交替の際に使用しているものなので、Web 会議中の話者交替においても自然に取り入れられると考える。また発話欲求のない場合には予備動作を抑えることも参加者にとって自然であると考えた。Web 会議において素早く頻繁に話者交替する会議を実施可能にするための新機能を学習し、使用するにあたり、上記の自然さが参加者への負担を軽減する大事な要素であると考えた。

- 3) 参加者自身により調整された発話欲求を伝達し、他の会議参加者に対して正確な判断材料となるシグナルの提示

それぞれの参加者の発話欲求がある閾値を超えたとき、参加者間で誰に発話欲求があるか否かを確認できるようにする。すなわち、発話欲求があると判定された参加者の映像を強調し、他の参加者へ示すようにする。本提案手法では、同時に強調される参加者の数に制限を設けず、自身が発話するか否かの判断は参加者へ委ねることとした。たとえばある参加者に発話欲求が生じた際に、他の参加者の映像が先に強調されていた場合、その参加者は自身の発話を控えて譲ることで発話衝突を回避することができる。

5. 発話欲求伝達モジュールの実装

会議システム全体の系としては、図 4 のようになる。商用 Web 会議システムの MeetingPlaza [14] と連携できるように発話欲求伝達モジュールを実装した。各クライアントマシンにおいて、発話欲求伝達モジュールはセンサから得られた情報を基に各参加者の予備動作候補を検知し、発話欲求フィードバックインジケータや、参加者映像へ黄色い枠を表示させる要求を MeetingPlaza クライアントモ

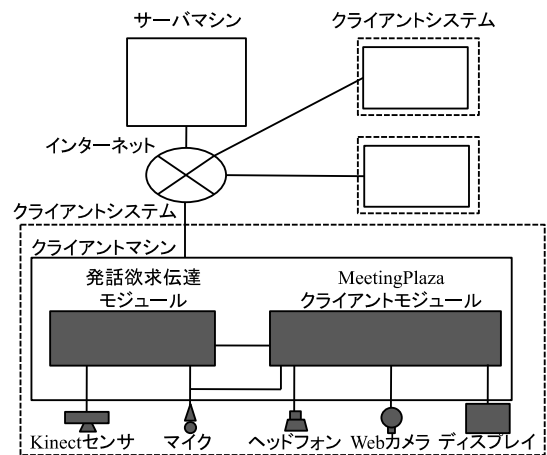


図 4 会議システム全体の構成図

Fig. 4 Structure of system.

ジュールへ渡す。MeetingPlaza クライアントモジュールはこれらの情報をインターネット経由でサーバマシンへ送る。サーバマシンは、受信した情報をすべてのクライアントマシンへブロードキャストする。以上の流れによってクライアントマシン間で情報が共有され、発話欲求フィードバックインジケータや黄色い枠の状態が同期される。Web 会議システム画面上に表示される映像は Web カメラから取得し、マイクから取得した音声とともに、MeetingPlaza クライアントモジュールへ取り込まれ、後は前述と同様の流れで、各クライアントマシンへ共有される。

発話欲求伝達モジュールは、予備動作候補検知部、発話欲求推定部、そして発話欲求伝達部からなる。

5.1 予備動作候補検知部

著者らはマジョリーの知見 [9] をもとに、「手を挙げる」、「手を顔周辺へ持っていく」、「傾く」、「身体を前後に動かす」が Web 会議中の発話前に行う予備動作であることを明らかにした [10]。このうち特に「手を挙げる」、「手を顔周辺へ持っていく」、「傾く」の 3 つが、参加者が意識的に行う予備動作であることが分かった。これらを用いることでシステムが発話欲求の推定をしやすく、また参加者自身が推定された発話欲求の度合いを調整しやすくなると考えた。参加者に発話欲求があるときに専用のボタンを能動的に押させる方法なども考えられるが、普段の会話で行っていない動作を用いることは、かえって参加者の負担になると考えた。

Kinect センサを用い、人を認識し、その頭部、首、肩関節、肘関節、手首などの三次元座標をリアルタイムに取得する（画像解像度 $1,280 \times 1,024$ 、奥行き方向解像度 640×480 ）（図 5）。本プロトタイプ実装時点では Kinect センサの画角に参加者の腰より上部が収まらなくては上記三次元座標を安定して取得できなかった。そのため Kinect センサをディスプレイ上部に取り付け、斜め下へ向けるこ

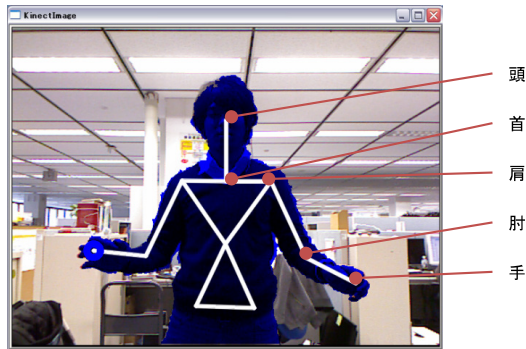


図 5 キャプチャされた参加者の関節の 3 次元位置: Kinect センサにより, 参加者の頭, 首, 肩, 肘と手の 3 次元位置座標が求められる

Fig. 5 Captured 3D positions of a participant's articulations: Kinect sensor detects 3D positions of participant's head, neck, shoulders and elbows.

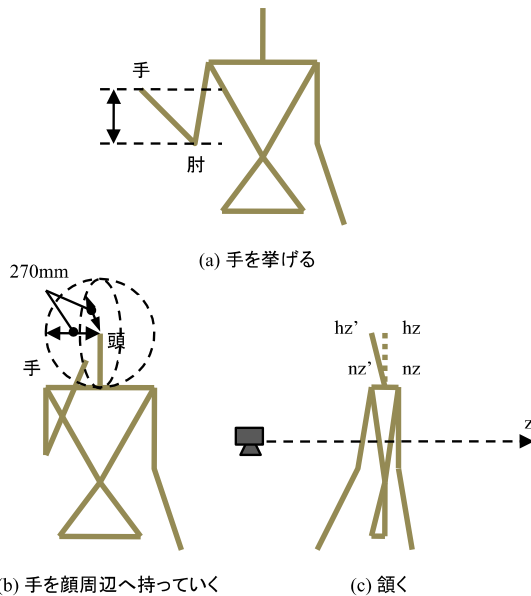


図 6 予備動作の検知方法

Fig. 6 Detection of prelim-motions.

とで参加者の腰より上部の RGB 画像, 奥行き画像を取得した. それぞれの予備動作候補の検知基準は以下のようにした.

〈手を挙げる〉

肘関節の位置よりも手首の位置が天地方向で高くなった場合, 手を挙げていると判定した (図 6(a)).

〈手を顔周辺へ持っていく〉

手首と頭部の三次元座標の距離が, ある一定の閾値 (今回は 270 mm に設定) よりも短くなった場合, 手を顔周辺へ持っていく予備動作候補であると判定した (図 6(b)).

〈頷く〉

奥行き方向の軸を z 軸とし, Kinect センサから離れる方向を正とする (図 6(c)). 頭部の z 座標値を hz , 首の z 座標値を nz , ある時間経過した後の z 座標値をそれぞれ hz' ,

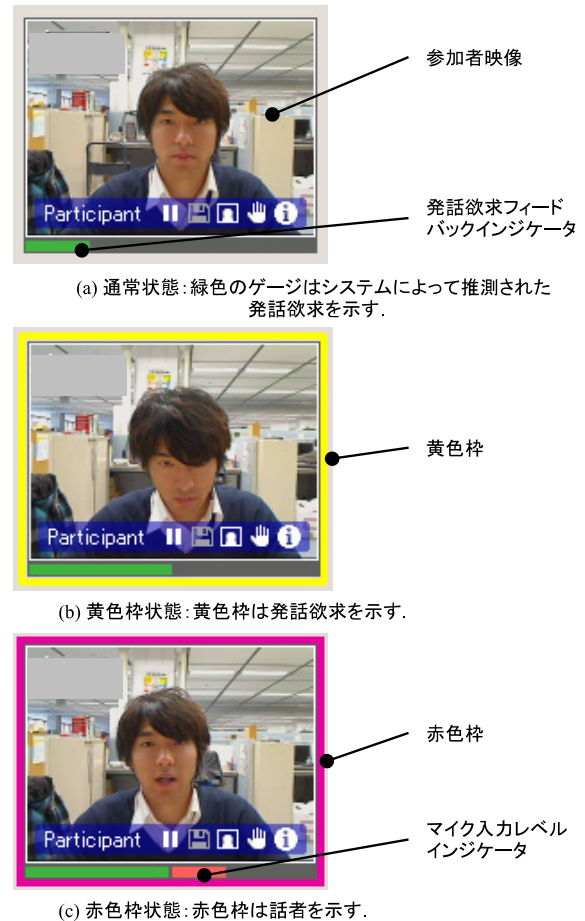


図 7 参加者映像

Fig. 7 Picture of participant.

nz' とすると,

$$hz - nz > hz' - nz' \quad (1)$$

が満たされる場合, 頷いていると判定した.

5.2 発話欲求推定部

前節で検出した予備動作候補を基に, 参加者ごとに発話欲求フィードバックインジケータを表示した (図 7(a)). 予備動作候補ごとに一定の値 (スコア) を決め, 検知された動作に対応したスコアの総和を発話欲求度とした. 発話欲求度は 0 以上の値をとるものとした. 発話欲求フィードバックインジケータは, 参加者映像下部左側に緑色のゲージとして表示させ, 同右側にはマイク入力音声レベルインジケータを赤色で表示させた (図 7(c)). 発話欲求度が 0 のとき, ゲージは表示されず, 100 のときにちょうど左半分を緑色のゲージが埋めるようにし, ゲージの長さで発話欲求度を表現した. 発話欲求度が 100 を超えた場合, ゲージは左半分を埋めた状態のままにした. また, マイクから入力音声がある場合, 発話欲求度はつねに 100 となるようにした. 発話欲求が 100 になり満たされてから発話にいたる (音圧が 1 以上の状態) までは連続的に見せることで, 参加者が「発話欲求のある状態」を自然に認知できると考

えた。発話欲求度は時間経過とともに一定の速度で減少させるようにした。

本手法において参加者にとって意味があるのは、発話欲求度が 50, 60 などの具体的な数値ではなく、100 に達しているか否か、100 に達しそうであるか否かである。参加者はこれを見て発話欲求の有無に応じて動きを制御し、発話欲求度を 100 まで上げるか否かを選択する。上記、予備動作候補ごとに設定されたスコアの大きさは、会議実施前に参加者が、予備動作をしてから発話することを数回繰り返す、微調整できるようにした。

5.3 発話欲求伝達部

ある参加者の発話欲求度が 100 を超えたときに、その参加者映像の枠を黄色く強調した (図 7(b))。提案機能を実装している商用システムの MeetingPlaza [14] は、発話中の参加者映像に赤色枠を表示する仕様となっている。発話の前段階の黄色から、発話中を赤色へ変化することは、人が普段慣れ親しんでいる信号機の色の変化と等しく、参加者が状態の遷移をイメージしやすくなると考えた。

6. 評価実験

6.1 実験概要

この実験の目的は、本プロトタイプにより、発話欲求を正しくシグナルへ反映し、伝達でき、発話衝突を低減させることができるかどうか、ユーザ評価実験にて確かめることである。具体的には、3 人の被験者に Web 会議システムを用いてアイデアを創造する会議を行い、発話欲求を伝達し、発話衝突確率が低減できるかを検証した。

比較条件として、通常の Web 会議システムを使用した条件と、発話欲求を伝えるために「ボタン」を使用した条件とを用意した。

6.2 実験方法

20 代男女 9 人の被験者を会議の参加者とした。被験者は 3 人 1 組のグループに分かれた。グループを構成する被験者は、互いに親しい間柄であった。遠隔会議システムには MeetingPlaza [14] を使用した。被験者はそれぞれ、防音室に入り、クライアント端末から MeetingPlaza にログインした。会議参加者の映像は、2 人分を横に並べて配置し、もう 1 人を下に配置した。すべての参加者は、マイクとヘッドホンの組み込まれたヘッドセットを装着した。いずれかの参加者のマイクに音声入力があった場合、その参加者映像に赤色の枠が表示されることで、その参加者が発話中であることが、すべての参加者へと伝えられた。

実験条件として下記 3 つを用意した。

- 1) Web 会議条件：上記話者を知らせる機能のみを使用し、参加者には特別な操作を要求しない。
- 2) ボタン制御条件：提案手法と比較するため、単純な操

表 1 被験者グループごとの実施した条件順序

Table 1 Order of conditions each subjects' group took.

	1 回目	2 回目	3 回目
グループ 1	条件 1	条件 2	条件 3
グループ 2	条件 2	条件 3	条件 1
グループ 3	条件 3	条件 1	条件 2

作で明確な発話欲求を伝えることができる仕組みのうち、最もシンプルだと考えられる条件を用意した。上記話者を知らせる機能に加え、参加者が発話したいときにキーボードのスペースキーを押下することで、その参加者の映像に、黄色枠を表示でき機能を持たせた。

- 3) 提案手法条件：上記話者を知らせる機能に加え、発話欲求伝達モジュールを使用し、発話欲求フィードバックインジケータと、発話欲求を他の参加者へ伝えるための黄色枠表示機能を持たせた。被験者には実験前に、検知される予備動作の種類を教示し、黄色枠を使用して発話欲求を伝えるよう指示した。図 1 のように、画面右側に、自身の映像と、予備動作が検知されているか否かをモニタリングできるウィンドウを表示した。

条件 2) と 3) では、被験者へ黄色枠の表示されるルールを説明し、「黄色枠を活用して話者交替を行ってください」と指示した。本実験では提案手法の機能を用いて会話することでどの程度発話衝突を低減させることができるかという点を評価することが目的であるため、黄色枠を活用するよう指示した。今後実用性評価をする段階においては、被験者へ黄色枠の表示されるルールのみを伝えて実験を行い、提案機能がどの程度利用されるかを調査することも必要であると考えた。上記 3 つの条件で、8 分間ずつ会話をし、操作に慣れてから実験に臨んだ。

会議の内容は、携帯電話の新機能のアイデア出しをすることであった。被験者は 30 分間議論を行い、10 分ごとに条件を変更した (表 1)。著者らはこれまでに類似の実験を行い、1 時間未満の会議において、時間経過とともに発話回数が減少しないことを確認している [6], [10]。本実験においても 3 通りの組合せで実験を行い、順序効果がないことを確認し、実験を終了した。

各条件実施直後に、被験者に「発話が衝突したと感じた度合い」を 5 段階で回答させた。「発話がよく衝突した」と感じた場合は 5、そうでない場合は 1 とし、5 段階のスコアをつける主観評価を行った。

計測用にクライアント端末を 1 台用意し、MeetingPlaza へログインし、実験の様子を、実験者 2 人が観察した。1 人の実験者が会話のターン数をカウントし、もう 1 人が発話衝突回数をカウントした。会議中の音声を聞きながら、話者が切り替わったときにターンが遷移したと判定し、また、2 人以上の参加者が同時に発話を開始したときに発話

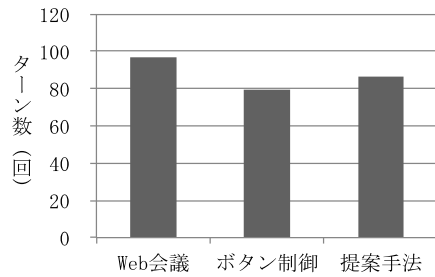


図 8 ターン数の比較：3つの条件で、ターン数の差は認められなかった

Fig. 8 Comparison of numbers of turns: There were no significant differences in numbers of turns.

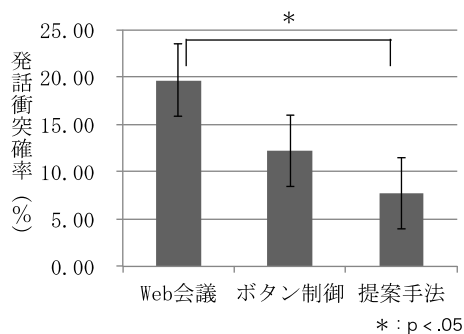


図 9 発話衝突確率の比較：提案手法条件では、Web 会議条件よりも、発話衝突確率が低かった

Fig. 9 Comparison of rates of speech contention: Speech contention rate in our proposal condition was lower than that in Web conference condition.

衝突が生じたと判定した。

6.3 実験結果

ある話者が発話開始してから発話終了するまでをターンと呼ぶこととし、実験条件ごとのターン数を図 8 に示す。Web 会議条件で、最もターン数が多く、次が提案手法条件となった。一元配置分散分析を行ったところ、群間自由度 2、群内自由度 6、P 値 0.869 であり、ターン数に統計的な差異は認められなかった。

条件ごとの発話衝突確率を図 9 に示す。本稿では、発話衝突回数数をターン数で割った値を発話衝突確率とした。発話衝突確率は Web 会議条件で最も高く、提案手法条件で最も低くなった。一元配置分散分析を行ったところ、群間自由度 2、群内自由度 6、P 値 0.099 であり、有意水準 10% で、条件により発話衝突確率が異なるという結果が得られた。Bonferroni-Holm の方法により有意水準を補正し、t 検定を用いて多重比較したところ、Web 会議条件よりも、提案手法条件の発話衝突確率の方が、有意水準 5% (補正後の有意水準は 1.67%) で有意に低いという結果が得られた。

続いて「発話が衝突したと感じた度合い」に関する主観評価の結果を図 10 に示す。Friedman 検定を実施したところ、自由度 2、カイ二乗値 8.07、有意確率 (P 値) 0.018 (< 0.05) となり、会議条件ごとに主観評価結果に差がある

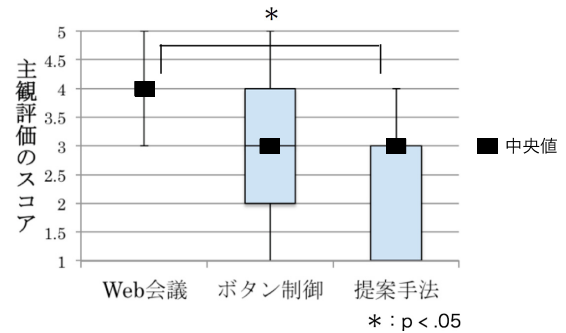


図 10 「発話が衝突したと感じた度合い」の比較：提案手法条件では、Web 会議条件よりも、発話が衝突していないと被験者が感じていた

Fig. 10 Comparison of how frequently subjects felt speech contentions occurred: They felt there were less speech contentions in our proposal condition than in Web conferences.

ことが分かった。そこでさらに Bonferroni-Holm の方法による有意水準の補正を行い、Wilcoxon の符号順位検定を用いて多重比較を行ったところ、Web 会議条件は提案手法条件と比較し、有意差 5% で ($p < .0167$) 「発話がよく衝突した」と被験者が感じたことが分かった。

6.4 考察

Web 会議条件、ボタン制御条件、提案手法条件のうち、提案手法条件での発話衝突確率が最も低かった。Web 会議条件と比較すると平均値が半分以下であり、統計的に有意な差がみられた。このことから、提案手法により、発話欲求を伝達して発話衝突を低減させることができたといえる。また主観評価の結果から、被験者が提案手法条件では Web 会議条件よりも発話衝突が少ないと感じたことが分かった。本実験結果ではボタン制御条件と提案手法条件の間に統計的差異は認められなかったが、定量的評価と主観評価両方において、提案手法条件で最も良好な結果が得られた。

ボタン制御条件は最も明示的かつ単純に発話欲求を伝達できる条件であったが、Web 会議条件と比較して、定量評価、主観評価結果ともに有意な差がみられなかった。録画したビデオを解析したところ、ボタン制御条件ではボタンを押して発話したにもかかわらず発話衝突する場面が多くみられた。ボタン制御条件ではボタンを押せばすぐに黄色枠が表示されるため、ボタンを押した後にすぐ発話することが多くみられた。一方「挙手」は自分のインジケータがたまるのを待つ間に、他の参加者に黄色枠が表示されるか否かを見て、確認してから発話する様子がみられた。ある参加者に発話欲求が生じた際に、他の参加者に黄色枠が表示されるか否かを確認する程度が異なることが、提案手法条件とボタン制御条件における発話衝突を低減させる効果の違いに現れた可能性が考えられる。

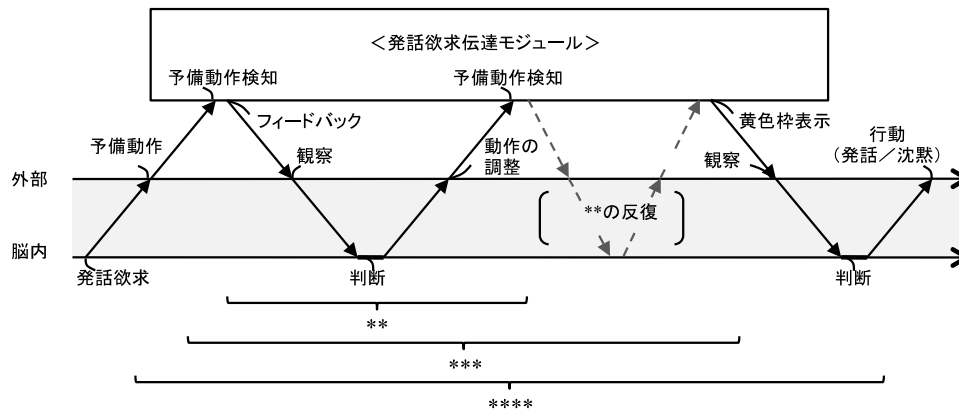


図 11 話者交替において提案手法を適用した際の認知行動モデル：***間のステップが図 2 のモデルへ加わった。参加者はフィードバックインジケータを観察し、自身の動作を調整する。**間のステップは、フィードバックインジケータと参加者の実際の発話欲求が合致するまで繰り返される

Fig. 11 Cognition and action model in turn-taking in our method: *** steps are added to the model in Fig. 2. Participants observe the feedback indicator and adjust their motions. ** steps repeat until the indicator fit their true desire to speak.

ビデオ解析によると、被験者は提案手法条件においてほとんどの発話の前に何らかの動作をし、発話欲求度を 100 にし、黄色枠を表示させてから発話する、という手順を踏んでいた。このことから被験者は自身に発話欲求があるときに、その発話欲求とシステムに検知されている発話欲求度との整合をとっていたことが分かる。また、黄色枠が表示されたことに気づくのが遅くて発話衝突する場面は観察されたが、黄色枠を無視して発話することはなかった。

提案手法条件やボタン制御条件では被験者に特別な操作を要求したため負荷がかかり、結果的に著しくターン数が減少するなどの可能性も考えられたが、実験結果に差は見られなかった (図 8)。しかしある被験者から「発話前に特別な動作を挟むことが自然な発話開始の弊害となる」というコメントが得られた。この点が、次に解決すべき課題であると考えられる。

話者交替において、本提案手法を用いた場合の認知・行動モデルは図 2 にいくつかのステップを追加し、図 11 のようなモデルになるという仮説を立てた。現状のプロトタイプでは、発話欲求が生じてから発話するまでに、1) 対面したコミュニケーション時よりも多くのステップを要すること、2) その操作を行う頻度が高いこと、が参加者へ負担をかけることが考えられる。これらの 2 点の要因による影響を低減させるために以下の項目を検討すべきであると考えた。

1) 最初の予備動作検知ステップと黄色枠表示ステップの間を短くするために、図 11 の**で示されたフィードバックステップと予備動作検知ステップ間の繰り返し回数を減らす。

1-a) 予備動作に割り当てるスコアを高く設定し、発話欲求フィードバックインジケータのゲージが溜まる速度を速くする。

1-b) 予備動作候補を検知した後に発話する回数が多いほど、その動作のスコアを高めていく。

2) 予備動作ステップと判断ステップ (図 11 の****) を通る機会を減らす。

2-a) すべての話者交替時に、予備動作をして発話欲求を伝達する手順を必要とせず、発話衝突のリスクがある場合のみ、参加者にその予備動作をして発話欲求を表すことを促す。

今後はこれらの方法を具体的に検討していき、発話衝突が少なく円滑な話者交替が可能で、かつ参加者への負荷が少ない Web 会議システムの実現を目指す。

7. おわりに

本稿では、手軽に使用できる Web 会議において、発話の衝突を低減させ、円滑な話者交替を可能にすることを目指した。参加者各々に発話欲求フィードバックインジケータを表示させ、参加者の動作からシステムが推定する発話欲求度合いをリアルタイムに表示し、参加者自身が動作を調整することによりその発話欲求度合いが設定された閾値を超えた場合、その参加者の映像に黄色い枠を表示し、他の参加者へと知らせる「発話欲求伝達手法」を提案した。

プロトタイプシステムを実装し、ユーザ評価実験を実施した結果、提案手法により発話欲求を他の参加者へ伝えられ、発話衝突を通常 Web 会議システムの半分に以下に低減させることができることが分かった。さらに被験者の主観評価でも、発話衝突が低減されたと感じたことが分かった。一方、意識的に予備動作を実施する頻度が高いことと、その一連の操作が自然な発話開始を阻害する可能性を持っていることも分かった。

今後は、低い発話衝突確率を保ったまま、より低負荷に、

会議参加者の発話欲求を伝達する手法の確立を目指して研究を進める。

参考文献

- [1] 2008 年版テレビ会議/Web 会議の最新市場と HD 化動向, シードプランニング (2008).
- [2] 高橋 誠: 会議の進め方, 日本経済新聞出版社 (1987).
- [3] 徳 勲, 友保康成, 渋谷 雄, 田村 博: テレビ会議技術の課題と利用法についての考察, *8th Symposium on Human Interface*, pp.207-212 (1992).
- [4] 鎧沢 勇, 滝川 啓, 大久保榮, 渡辺義郎: 衛星通信を利用した画像会議におけるエコー及び伝搬遅延の影響, 電子通信学会論文誌 (B), Vol.J64-B, No.11, pp.1281-1288 (1981).
- [5] Sacks, H., Schegloff, A.E. and Jefferson, G.: A Simplest Systematics for the Organization of Turn-Taking for Conversation, *Language*, 50, 4Pt1, pp.696-735 (1974).
- [6] 玉木秀和, 中茂睦裕, 東野 豪, 小林 稔: 人のコミュニケーションリズムに着目した Web 会議円滑化手法, IEICE Technical Report MVE2009, pp.101-106 (2009).
- [7] 玉木秀和, 東野 豪, 小林 稔, 井原雅行: 遠隔会議における発話の衝突と精神的ストレスの関係, グループウェアとネットワークサービス研究会, Vol.2011-GN-79, No.10, pp.1-6 (2011).
- [8] 松尾 隆: コミュニケーションの心理学, ナカニシヤ出版 (1999).
- [9] マジョリー・F・ウォーガズ: 非言語コミュニケーション, 新潮社 (1987).
- [10] 玉木秀和, 東野 豪, 小林 稔, 井原雅行: 遠隔会議における発話衝突低減手法, 情報処理学会論文誌, Vol.53, No.7, pp.1797-1806 (2012).
- [11] Sellen, A.J.: SPEECH PATTERNS IN VIDEO-MEDIATED CONVERSATIONS, *CHI'92*, pp.49-59 (1992).
- [12] シスコシステムズ合同会社: TelePresence, 入手先 (<http://www.cisco.com/en/US/products/ps7060/index.html>) (参照 2012-04).
- [13] Inkpen, K., Hegde, R., Czerwinski, M. and Zhang, Z.: Exploring Spatialized Audio & Video for Distributed Conversations, *CSCW 2010*, pp.95-98 (2010).
- [14] NTT アイティ株式会社: MeetingPlaza, 入手先 (<http://www.meetingplaza.com/index-j.html>) (参照 2012-04).
- [15] 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 (2008).
- [16] Kawashima, H., Nishikawa, T. and Matsuyama, T.: Visual Filler: Facilitating Smooth Turn-Taking in Video Conferencing with Transmission Delay, *CHI2008 Extended Abstract*, pp.3585-3590 (2008).
- [17] Cassell, J. and Vilhjálmsdóttir, H.: Fully Embodied Conversational Avatars: Making Communicative Behaviors Autonomous, *Autonomous Agents and Multi-Agent Systems*, Vol.2, pp.45-64 (1999).
- [18] López, B., Hernández, Á., Díaz, D., Fernández, R. and Hernández, L.: Design and validation of ECA gestures to improve dialogue system robustness, *Proc. Workshop on Embodied Language Processing*, pp.67-74 (2007).
- [19] Wu, W., Arefin, A., Kurillo, G., Agarwal, P., Nahrstedt, K. and Bajcsy, R.: Color-plus-Detail in 3D Tele-immersive Video: A Psychophysical Approach, *ACM MM'11*, pp.13-22 (2011).
- [20] Xbox Kinect センサ, 入手先 (<http://www.xbox.com/ja-JP/kinect>) (参照 2012-04).



勤務。

玉木 秀和 (正会員)

2008 年慶應義塾大学大学院理工学研究科修士課程修了。同年日本電信電話(株)入社。NTT サイバーソリューション研究所にて CSCW, ヒューマンインタフェースの研究に従事。現在, NTT コミュニケーションズ (株)



ンタフェース, 障がい者メディアの研究に従事。

東野 豪 (正会員)

1988 年東京工業大学大学院総理工学研究科物理情報システム専攻修士課程修了。同年日本電信電話株式会社入社。現在, NTT サービスエボリューション研究所主幹研究員。画像符号化, 映像配信システム, ヒューマンイ



ティ, マルチモーダルインタフェース等の研究に従事。現在, サービスエボリューション研究所主幹研究員。ACM, IEEE Computer Society, 電子情報通信学会, 画像電子学会, 日本バーチャルリアリティ学会, 可視化情報学会各会員。博士 (工学)。

小林 稔 (正会員)

1988 年慶應義塾大学理工学部卒業, 1990 年同大学大学院修士課程修了。同年日本電信電話(株)入社, 1996 年マサチューセッツ工科大学修士課程修了。映像を用いたコミュニケーションメディア, バーチャルリアリ



にて客員研究員。2010 年東京工業大学大学院博士課程修了。現在, NTT サービスエボリューション研究所主幹研究員。ACM, 電子情報通信学会, 画像電子学会各会員。工学博士。

井原 雅行 (正会員)

1994 年東京工業大学大学院修士課程修了。同年日本電信電話株式会社入社。仮想空間共有コミュニケーション, 価値観共有の研究等に従事。2002~2003 年加国 New Media Innovation Center およびブリティッシュコロンビア大学