情報処理学会論文誌 Vol.53 No.7 1797–1806 (July 2012)

遠隔会議における発話衝突低減手法

玉木 秀和[1,a)] 東野 豪[1] 小林 稔[1] 井原 雅行[1] 岡田 謙一[2]

受付日 2011年10月24日, 採録日 2012年4月2日

概要：Web 会議システムは利用や導入の手軽さがあるが，個々の参加者の映像が小さく，映像品質に制限があるため，誰がいつ発話し始めるのかを判断しにくく，発話が衝突してしまうことが多い．このため，発話の意欲が低下し，時間効率が悪くなり，生産性の低い会議になりかねない．人は普段，対面したコミュニケーションではノンバーバル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が，Web 会議ではこれを行うことが難しい．そこで本研究では，Web 会議において，人が発話の前に行う特徴的な動作を検知して最も次に発話しそうな参加者を決定し，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する．提案概念を実現するプロトタイプを実装し，会話実験を行ったところ，システムが検出する特徴的な動作を事前に教示する場合は，本提案手法を用い，発話衝突確率を低減させられることが分かった．

キーワード：遠隔会議，話者交替，発話衝突

Method of Reducing Speech Contention in Distributed Conferences

Hidekazu Tamaki[1,a)] Suguru Higashino[1] Minoru Kobayashi[1]
Masayuki Ihara[1] Ken-ichi Okada[2]

Received: October 24, 2011, Accepted: April 2, 2012

Abstract: In Web conferences, we sometimes cannot recognize when other participants begin speaking. This depresses the participants' motivation and wastes time. We can take turns smoothly in face-to-face communication through the use of non-verbal messages, but this back channel is not available in existing web conference systems. We overcome this problem by proposing a method that senses actions that indicate the desire to speak and visualize them to the other participants. An evaluation of a prototype indicates that it realizes smoother turn-taking than is possible in usual web conference systems, when participants were noticed which motions the system detected.

Keywords: distributed conferences, turn taking, speech contention

1. はじめに

環境負荷制約の高まり，世界不況，パンデミック対策などの影響により遠隔会議システムの需要が増加してきている．遠隔会議システムの中でも Web 会議システムは，自席で利用できるうえ，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]．

しかし，Web 会議は TV 会議専用システムに比べて，デスクトップ上で利用し，転送効率が保障されないネットワークを通じて行うという特性上，個々の参加者の映像が小さい，フレームレートが低い，画質が悪い，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限がある．このため他の参加者の様子を読み取りにくく [2]，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，ということが多い．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 [3]．また，Sacks らは発話が衝突すると発話を諦めて中断する傾向が高いことを指摘している [4]．このような状態では，素早い話者交替

[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
NTT Cyber Solutions Laboratories, NIPPON TELEGRAPH AND TELEPHON CORPORATION, Yokosuka, Kanagawa 239–0847, Japan
[2] 慶應義塾大学理工学部情報工学科
Department of Computer and Information Science, Faculty of Science and Technology, Keio University, Yokohama, Kanagawa 223–8522, Japan
[a)] higashino.suguru@lab.ntt.co.jp

を頻繁に行う会議はしにくく [3]，消極的な会議になりかねない．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 [5]（アイデアを考え出す会議）を行う場合に，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]．

人は普段，対面したコミュニケーションの場面では頷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 [7], [8]．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し，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．そのために著者らは，Web 会議において，人が発話の前に行う特徴的な動作を検知して，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する．

本稿ではまず，本研究の位置づけを示し，続いて Web 会議において発話の衝突が起こっている状況を整理する．そして人が発話しようとしたときに行う特徴的な動作が，Web 会議中の会話でどの程度活用されているのかを分析する．これらをふまえ，発話衝突を低減させる提案コンセプトを説明し，これを実現するプロトタイプシステムの構築と，その効果を評価するためのユーザ評価実験の結果と考察を述べる．

2. 関連研究

2.1 対面コミュニケーションにおける話者交替

Sacks らが見出した話者交替ルール [4] によると，現話者が次話者を指定せずに発話を終えた際には，次に最も早く発話した者が次話者となるとされている．この場合，2 人以上の参加者が同時に発話を開始し，発話が衝突してしまう恐れがある．マジョリーは，ある参加者が次の発話権を獲得するためには非言語情報を活用すると述べ，いくつかの典型的な動作をあげている [8]．それは腕組みをほどく，身体を前に乗り出す，現話者の方へ向き直る，現話者と視線を交わす，目立つように頷くといった動作である．著者らは本稿において，上記のように，次話者が発話しようとしたときに行う特徴的な動作のことを「予備動作」と呼ぶこととする．なおこの予備動作は，あくまで発話意図があるときに行われる動作を指し，その動作の後，実際に発話に至らなかったものも含めることとする．

2.2 話者交替支援システム

様々な映像コミュニケーションシステムが開発され，伝達できる情報量が増加している．しかし，Sellen はこれらにはまだコミュニケーションを行う際に弊害があると指摘した [9]．1 対 1 の遠隔コミュニケーションであっても，遅延のある環境では発話の衝突が多く起こる [3]．1 対 1 のコミュニケーションであれば，発話権が交互に遷移するが，多人数のコミュニケーションではそれは複雑になる．商用化されている会議システムにも発話権の遷移をサポートする機能が実装されているものがある．ボタンを押してマイクのオン・オフを切り替えることにより発話権を獲得，放棄する機能や，挙手アイコンを表示させるためのボタンを備えたものもある [10]．

人と機械の会話を自然に感じさせるために発話前の非言語情報を表現する機能をエージェントシステムに組み込んだ試みも存在している [11], [12]．

石井らはアバタを介した遠隔コミュニケーションシステムにおいて，次に発話することを促す視線を表現する手法を提案した [13]．しかしこの手法では，音声を基に動作を疑似的に作り出しており，実際の参加者の動作は反映されない．

Kawashima らは 2 者間の遠隔コミュニケーションにおいて，フィラー（間を埋める音声）を，映像によって表現した [14]．しかし参加者は 2 人に限定され，これも実際の参加者の動作は反映されない．

そこで著者らは，実際の参加者の予備動作をセンシングすることで得た情報を基に，他の参加者へ何らかの情報提示をすることで，発話の衝突を解消する仕組みの構築を目指す．

3. 分析

Web 会議における発話の衝突を低減させる手法を見出すために，(1) Web 会議における発話衝突のメカニズムと，(2) Web 会議における予備動作の使われ方を分析した．

3.1 Web 会議における発話衝突のメカニズム

商用 Web 会議システム MeetingPlaza [10] を用い，Web 会議システムの使用経験がない 4 人の被験者で 10 分間の会議を実施させ，その様子をビデオに記録した．会議の議題は，4 人が参加するキャンプで実施するレクリエーションを考えることとし，司会者を設けずに会議を行わせた．予備動作の使用頻度，種類，効果を参加者間でまんべんなく調べるため，記録したビデオのうち参加者間で発話頻度に偏りの少ない 5 分間を分析の対象とした．対象とした区間において発話と予備動作，発話の衝突を記録した．なお，意図的な衝突は本研究の対象から外れるのでこの分析からは省いた．意図的な衝突とは，他の参加者の発話開始後，その発話と重なることを分かったうえで発話開始することにより生じる衝突を指す．

この会議において観察された予備動作は以下の 4 つであった．

- 手：口元や顔周辺へ持っていく動作
- 頭：横へ動かす動作
- 頷く

- 音声：相槌や笑いなど，発話に対するポジティブなフィードバックを返すもの．

上記 5 分間に 45 回の話者交替が起こり，11 回の発話衝突が観測された．発話の衝突は，次の 2 つの状況に分類できた．

1) 複数の参加者が同時に発話して衝突
2) 発話の切れ目（呼気段落）に割り込もうとして衝突

実際に発話の衝突が起こったシーンの具体例を 2 つあげて解説する．

＜複数人の参加者が同時に発話して衝突＞

1 つ目の例を**図 1** に示す．記録映像のタイムスタンプを横軸とし，参加者ごとに発話や行っている予備動作を記述した．まず参加者 1 が全体に投げかけるような発話をして終了した．参加者 2〜4 は発話の衝突が起こる 10 秒前より各々，予備動作である可能性のある動作をしていた．参加者 2 はさらに 3 秒後にもう 1 度頭を動かす予備動作を行っていた．しかし後のヒアリングによると，このとき参加者 2〜4 は参加者 1 の映像を見ていて，互いの動作は認知できていなかった．結果的に参加者 2 と 3 が同時に発話開始し，衝突してしまった．参加者 3 は発話を中断し，参加者 2 は発話を継続した．

＜発話の切れ目に割り込もうとして衝突＞

2 つ目の例を**図 2** に示す．参加者 1 が発話をしていた．参加者 3 は，参加者 1 の発話音声が途切れた瞬間を狙って，発話を開始した．発話の前には，頭を横に動かす予備動作をしていたが，後のヒアリングによると，参加者 1 はそれに気付かず発話を継続した．結果的に発話は衝突し，参加者 3 は発話を中断，参加者 1 は発話を継続した．

3.2 Web 会議における予備動作の使われ方の分析

3.2.1 手順

前節の発話衝突場面の例では，いずれも予備動作を行ってから発話したにもかかわらず，発話の衝突が起こっていた．これらの予備動作の活用度合いが Web 会議と対面会議でどの程度異なるかを確かめるため，前節で説明したものと同様の会議を，対面会議で実施し比較した．対面会議においても発話，予備動作，発話の衝突を記録した．

ある参加者が予備動作を行ってから発話するまでの間に，別の参加者の発話が挿入されることは十分考えられる．しかし現状，予備動作をしたときの発話意図が，いつまで持続したかを測る手段はない．そのため，今回の分析では，予備動作である可能性のある動作を行った参加者が，すぐ次のターンに発話権を獲得し，発話を開始したか否かの関係のみに着目し，集計を行った．

3.2.2 予備動作の出現回数と効果

対面会議では 5 分間で 85 回の話者交替が起こり，1 回の発話衝突が観測された．**表 1** は，対面会議と Web 会議において予備動作である可能性のある動作が観察された回数と，その動作を行った後に発話した回数，そして後者を前者で割った値を動作後の発話確率として比較し，まとめたものである．Web 会議では，この予備動作である可能性のある動作の合計回数は 78 回であり，対面会議での 195 回と比較して半分以下であった．

この，予備動作である可能性のある動作の回数には，発

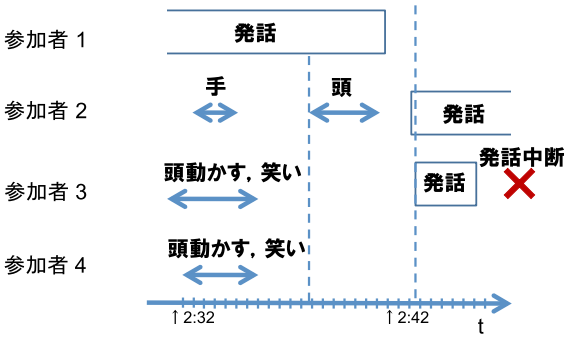

図 1 複数の参加者が同時に発話して衝突する様子
Fig. 1 Contention when a participant tried to speak during a speech pause.

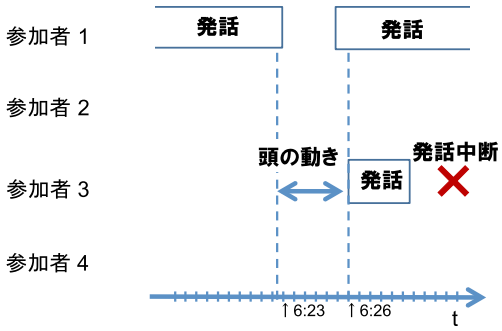

図 2 発話の切れ目に割り込もうとして衝突する様子
Fig. 2 Contention when some participants try to start speaking simultaneously.

表 1 動作後の発話確率
Table 1 Rate of speaking after “prelim-motions” Web.

予備動作である可能性のある動作	対面会議での発話確率（発話回数／動作回数）	Web 会議での発話確率（発話回数／動作回数）
手	61%(28/46)	73%(8/11)
頭	55%(41/75)	43%(15/35)
頷き	57%(17/30)	71%(5/7)
音声	57%(25/44)	60%(15/25)
合計	57%(111/195)	55%(43/78)

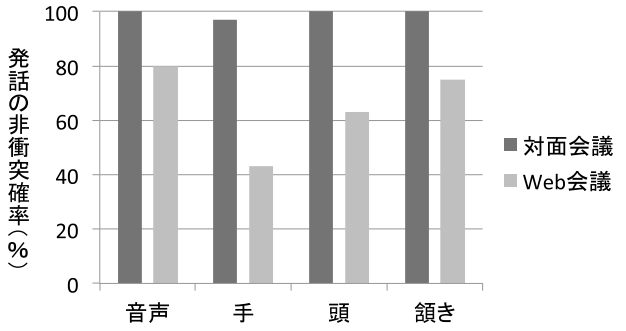

図 3 予備動作後の発話の非衝突確率
Fig. 3 Success rates of utterances.

話意図をともなう予備動作の回数と，発話意図をともなわない，その他の動作の回数が含まれる．現状，予備動作とその他の動作を見分けることは容易ではない．しかし，ある参加者が発話をした場合，その直前に行っていた動作は，予備動作であるといえる．したがって，予備動作である可能性のある動作をした後に発話をし，かつそれが衝突しなかった確率を求めることで，予備動作後の発話の非衝突確率を求めることができる（**図 3**）．対面会議では予備動作後の発話はほぼ 100%に近い確率で成功しているが，Web 会議ではどれも 80%を切っていた．表 1 から，予備動作全体に占める音声の割合は対面会議で 23%（44 回/195 回）であるのに対し，Web 会議では 32%（25 回/78 回）と多かったことが分かる．しかしその予備動作の可能性のある動作後でも 20%の発話が衝突している．手や頭の動きによる予備動作後の発話の非衝突確率は 40%，60%と低く，このことから，Web 会議では映像チャネルを通じて伝達される予備動作は認知しにくいことが推測できる．

以上をまとめると，今回の分析範囲内では，Web 会議では対面会議と比較して予備動作である可能性のある動作の回数が半数以下に減少し，予備動作後の発話の非衝突確率も大きく減少する傾向があった．

3.2.3 予備動作出現の特徴

今回の分析範囲内では予備動作表出の特徴として以下の傾向が見られた．

1) 予備動作の種類により，その後の発話確率が異なった（表 1）．
2) 最も多くの予備動作をした参加者が発話する割合が高かった．

図 4 に，Web 会議での 45 回の話者交替のうち，直前の発話中に行った予備動作多さの順位と発話の関係を示す．今回の分析では，半分以上の話者交替時に，最も多くの予備動作を行った参加者が発話していた．なお，最も多く予備動作を行った参加者が発話せず，それ以外の参加者が発話した 7 回のうち 2 回は，呼びかけに反応するなど，明らかに次話者を指定された場面であった．

3) 予備動作の種類によって，出現してから発話開始するまでの間隔が異なった．

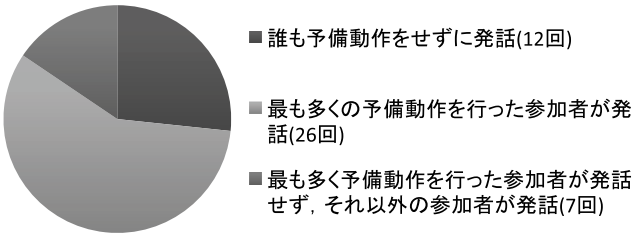

図 4 Web 会議における予備動作回数と発話の関係
Fig. 4 Relation between numbers of “prelim-motions” and utterances in web conferences.

手と頭による予備動作は，表出後すぐに発話することが多く観察された．また音声による予備動作も，その予備動作後そのまま発話に移ることが多かった．それらに対して頷きは，直前の他者の発話の中ごろから表出し，その発話が終了してから発話することが多く観察された．

4. 提案

4.1 次話者候補提示

3.2.2 項の分析結果は，普段我々が対面コミュニケーションで話者交替を円滑にするために活用している予備動作が，Web 会議では認知されていない可能性を示唆していた．そこで著者らは，予備動作を検知し，そこから最も次に発話しそうな参加者を選定し，それをすべての参加者へ提示することによって，発話の衝突確率を低減させる手法を提案する．この手法の実施イメージは以下のとおりである．ある参加者が発話しようとして予備動作をすると，その動作が検知され，次の話者候補（次話者候補）として選定される．この参加者の映像が強調表示され，他の参加者からの注意が集まり，自然に発話を開始することができる．

4.2 予備実験

前節で述べた提案手法により，Web 会議における発話の衝突確率を低減させることができる見込みがあるかどうか，「オズの魔法使い実験」により確認した．オズの魔法使い実験とは，評価対象となるシステムのうち，実装できていない部分を，裏に隠れた人が代替して操作することでシステム全体の有効性を検証する実験である [15]．

4.1 節で説明した提案手法のうち，参加者の予備動作を検知し，次話者候補を選定する部分を，人の手によって行う，オズの魔法使いシステムを構築した．魔法使い役（実験者）は，Web 会議映像・音声をリアルタイムに視聴し，予備動作を観察した．そして，3.2.3 項に述べた予備動作出現の特徴を基に次話者候補を選定した．ただし，3.2.3 項 1) と 3) に関しては，魔法使い役の実験者の経験的な感覚を頼りにするものとした．あらかじめ参加者ごとに ID を振っておき，その ID に対応するキーをタイプすると，相当する参加者映像に青い枠が表示されるようにシステムを実装した．

表 2 発話の予備動作のスコア付け
Table 2 Score of “prelim-motions”.

予備動作	予備動作後の発話確率	スコア	有効時間
手	73%	8	1 秒
頭	43%	4	1 秒
頷き	71%	7	3 秒
音声	60%	6	音声入力のある間

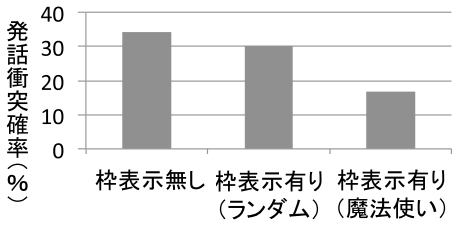

図 5 オズの魔法使い実験の結果
Fig. 5 Result of “Wizard of Oz” test.

上記オズの魔法使い条件と比較するために，システムにより 5 秒に 1 度，ランダムに次話者候補を選定し，その参加者映像に青い枠を表示させるシステムを構築した．以上 2 条件に，従来どおりの Web 会議システムを用いた，枠表示なし条件を加えた 3 条件にて比較実験した．4 人の参加者で，それぞれの条件につき，5 分間の会議を実施した．議題は，「テレビの新機能を考える」とした．実験を行った順序は，「枠表示なし」「枠表示あり（魔法使い）」「枠表示あり（ランダム）」とし，後 2 つの条件ではどちらも「次に発話しそうな参加者に枠が表示される」と教示した．

実験結果は**図 5** のとおり，枠表示あり（魔法使い）条件で発話衝突確率を下げることができた．ランダムに枠を表示した条件では発話衝突確率を減らすことができなかった．

5. プロトタイプ実装

提案概念を実現する第 1 のプロトタイプを実装した．3 章での分析を基に，予備動作を検出し，「次話者候補」をすべての参加者に示す機能を実装し，MeetingPlaza に組み込んだ．

5.1 予備動作の検知

Web 会議システムで使用される Web カメラ，ヘッドセットから入力される映像，音声を基に以下の予備動作をリアルタイムに検知する機能を実装した．

- 手：手を上下に動かす動き
- 頭：頭が上下左右へ移動する動き
- 頷き：顔を縦に動かす動作
- 音声：話者でないときの発声

各動作の精密な音声・画像処理技術自体は本実験の対象外である．今回は簡単な認識機能を実装した．そのため Web カメラの正面に参加者が 1 人座り，背景が変化しない環境下で実装した．顔認識には OpenCV に組み込まれた認識機能を使用した．なお，プロトタイプシステムにおける参加者映像の解像度は 320 × 256，リフレッシュレートは 30 fps であった．

＜手＞

映像を縦に 4 分割したうち，左右両端いずれかの領域で，フレーム間差分をとり，手変化が生じた場合，手の動きと判定させた．

＜頭＞

顔領域の中心点の x，y 座標のうちいずれかが，3 フレームの間に 30 ピクセル以上移動した場合，頭の動きと判定させた．

＜頷き＞

顔領域の中心点がフレーム間で x 座標 ±2 ピクセル以内，かつ y 座標 ±30 ピクセル以内で移動した場合，頷きと判定させた．

＜音声＞

最大 1 人の参加者が話者として指定される．いずれかの参加者のマイクから音声が入力されている場合，最も先に音声入力のあった参加者が話者として指定される．話者の音声入力がなくなった場合，次に最も先に音声入力のあった参加者が話者として指定される．話者でないときの音声入力を，すべて音声による予備動作と判定させた．

5.2 次話者候補の選定方法

普段我々が対面コミュニケーションをする際には，予備動作を認知し，現発話とのタイミング，いくつかの動作の組合せ，などを複合的に処理して次話者を判断し，発話権授受を行っていると考えられている [16]．普段我々が対面コミュニケーションをする際には，予備動作を認知し [8]，現在の発話の切れ目を予測し，誰が次の話し手になるのかという次話者の選択も行っている [16] と考えられている．本稿では 3.2.3 項にまとめた予備動作出現の特徴を基に，シンプルな次話者候補選定アルゴリズムを組み，提案手法の有効性を探る．下記のとおり次話者の候補者が最も高い値（発話可能性ポイント）をとる方法を考案した．

(1) 予備動作のスコア設定

予備動作ごとに，その動作後の発話確率に応じたスコアと有効時間を付与（**表 2** 参照）した．発話したいという意図（発話意図）があったとしても，多人数会話であれば必ずしも発話権を獲得できるわけではない．そのため，今回の分析で得られた予備動作後の発話確率よりも，発話意図

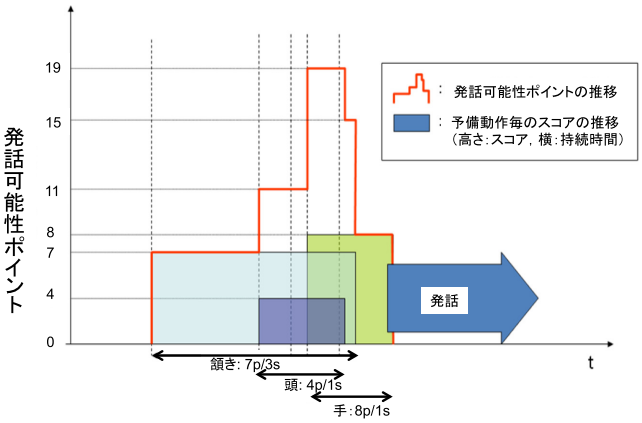

図 6 予備動作ごとのスコアの推移と発話可能性ポイント
Fig. 6 Calculation of the score.

は高いことが予想される．しかし今回の実装において著者らは，予備動作後の発話確率が高いほど，その動作をしたときの発話意図が高いものと仮定し，スコアを定めた．ここで，予備動作ごとに異なる，できるだけ小さな整数をスコアとして設定するために，予備動作後の発話確率を 9 で割り，端数を切り捨てた値を採用した．

(2) 有効時間

表出後すぐに発話にいたることの多い手と頭による予備動作は，有効時間を 1 秒に設定した．音声による予備動作は，その直後に発話に移ることが多かったため，音声入力のある間を有効時間とした．分析対象とした会議では 6 秒間程度の発話が多く観察されたため，発話する直前の発話の中ごろに表出する頷きの有効時間は，その半分の 3 秒とした．

(3) 発話可能性ポイント

ある時刻における発話可能性ポイントは，有効時間中にあるスコアの線形和，すなわち，ある人 n の時刻 t における発話可能性ポイント (n,t) は次式のように算出する．

$$\text{発話可能性ポイント } n(t) = \Sigma\text{スコア } n(j) \times \text{持続時間関数 } n(t-t0,j) \quad (1)$$

ここで，j は予備動作を表す識別子で，表 2 によれば，たとえば，1：手，2：頭，3：頷き，4：音声である．持続時間関数 $(t-t0,j)$ は，予備動作 j が検知された時刻 $t0$ から時刻 t が持続時間内であれば 1 を，時間外であれば 0 をとる関数である．これを**図 6** に示す．より正確にはスコアは持続時間内で刻々とリニアに変化し，さらに発話可能性ポイントは単純な線形和ではなく，いくつかの動作や発話とのタイミングにより左右するであろうが，まずは上記のような計算方法を実装した．ある時刻での発話可能性ポイントが最も高い参加者を次話者候補として選択した．

5.3 次話者候補の提示方法

上記基準により選択された次話者候補の参加者映像に黄色い枠（以降，黄色枠と表記）をつけることで，全参加者へと次話者候補であることを示した．MeetingPlaza には元々の機能として，話者に赤色の枠がつく仕様となっている．黄色から赤への遷移が我々にとって信号機でなじみ深く認知しやすいと考え，話者の前段階である次話者候補を表す色にも黄色を採用した（**図 7**）．

図 7 プロトタイプシステム実行画面
Fig. 7 Prototype system.

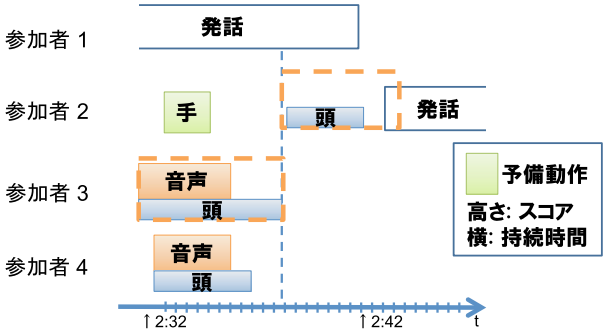

図 8 発話可能性ポイントの推移と次話者候補選択の例
Fig. 8 Image of deciding the nominee.

動作の例として，3.1 節であげた発話の衝突場面（図 1）に，本手法を適用していた場合の動作を説明する（**図 8**）．参加者 1 が発話している間，他の参加者 2〜4 にはそれぞれ行った予備動作に応じて，発話可能性ポイントが計算されている．ここで示されているタイムラインの前半部では参加者 3 の予備動作が多く，最も発話可能性ポイントが高いため，次話者候補を示すために，参加者 3 の映像に黄色の枠が表示されている．ところがタイムラインでの中盤になると，参加者 2 が頭を動かしたことによってそのスコアが足しあわされ，この時点で最も発話可能性ポイントが高くなった．すると今度は参加者 2 の映像に黄色の枠が表示される．そして参加者 1 の発話が終了した際に，参加者 3 と 4 は，枠に色がついていた参加者 2 に発話権を譲り，参加者 2 はそのまま発話を成功させることができる．著者らが MeetingPlaza を用いて往復の音声遅延を測定したところ 600 msec であったが，3 章の分析より，予備動作はこの遅延を加味しても十分前もって出現するため，この手法は発話の衝突を回避するために有効に機能すると考える．

6. 実験

6.1 目的

プロトタイプシステムを用いて，本提案コンセプトの有効性を検証する．具体的には，次話者候補の参加者映像に黄色枠を表示することで発話の衝突確率を低減させることができるかを確かめる．

6.2 手順

プロトタイプシステム（次話者候補を示す黄色枠あり）と，従来の Web 会議システム（黄色枠なし）との比較実験を行った．Web 会議システムには MeetingPlaza を使用した．それぞれのシステムを用いて被験者に創造会議を実施させ，発話の衝突確率を比較した．20 代〜30 代の男女 30 人の一般被験者を募り，3 人ずつの 10 グループを構成し実験を実施した．被験者のグループメンバ 3 人は互いに親しい知り合いであることを条件とした．会議の議題は「テレビの新機能を考える」とした．グループで話し合ってアイデアを出していき，グループ内に 1 人定めた書記係に，そのアイデアを記録させた．提案手法である黄色枠あり条件と，従来手法である黄色枠なし条件の 2 つを交互に行い，その順序はランダムとした．被験者 10 グループをさらに 5 グループずつの 2 群に分け，以下のとおり，群ごとに黄色枠あり条件において教示する情報を変えた．片方の 5 グループでは，「黄色枠は次に話しそうな参加者映像に表示される」とだけ教示し，もう片方の 5 グループには，上記内容に加えてシステムが検知している予備動作を具体的に教示した．前 5 グループは「動作非教示群」と記述し，後 5 グループは「動作教示群」と記述する．

被験者の疲労を考慮し，1 つのグループあたりの会議実施時間は合計 14 分間とした．前後半 7 分に分け，黄色枠あり，なしの条件をそれぞれ実施した．条件の順序はランダムに設定した．14 分間を通して議論は継続させた．いずれの条件においても，映像のリフレッシュレートは 30 fps に設定し，ある参加者が発声してから他の参加者へ伝わるまでの往復の音声遅延量は 400 ms 程度であった．実験中の会議システムの映像と音声をビデオカメラで記録し，解析を行った．

教示する内容は統一することが望ましかったが，どちらか一方を合理的に選択することができなかった．動作非教示群では，検知される動作を知らないユーザが自然にシステムを使用し，黄色枠に誘導され，話者交替を行うという状況である．自身の振舞いを意識せず，自然に会議をするため，発話回数は黄色枠なし条件と同等となり，発話衝突確率が減少することを予想した．一方動作教示群では，このシステムを商用化する際に説明書に記載される程度の内容，すなわち頷き，手，頭，音声といった具体的な予備動作の種類を教示してある．このため，積極的に予備動作を使用するようになり，黄色枠なし条件と比較して，発話衝突確率が減少するが，意図的な動作を挟む機会が増え，発話回数は減少することを予想した．

6.3 結果

それぞれの条件ごとに 7 分間の発話回数，発話衝突回数を記録した．また，発話衝突回数を，発話回数と発話衝突回数の和で割った値を，発話衝突確率として求めた．発話回数について，黄色枠なし条件（従来手法）を基準としたときの，黄色枠あり条件（提案手法）での増減を，動作非教示群に関しては**図 9** に，動作教示群に関しては**図 10** にまとめた．

動作非教示群では，5 つのグループすべてが，黄色枠あり条件において，発話回数が減少していた．一方，動作教示群では，黄色枠なし，ありの条件による増加，減少の傾向は見られなかった．黄色枠なし/あり条件間の発話回数の変化量の平均値を，動作教示/非教示群ごとに算出し**図 11** にまとめた．平均値の比較を t 検定で行ったところ，有意差 5%で，動作非教示群における黄色枠有無での発話回数の減少は，動作教示群のそれよりも大きくなることが分かった．

発話衝突確率についても同様に，黄色枠なし条件を基準としたときの，黄色枠あり条件での増減を，動作非教示群に関しては**図 12** に，動作教示群に関しては**図 13** にまとめた．動作非教示群では，黄色枠なし，ありの条件による

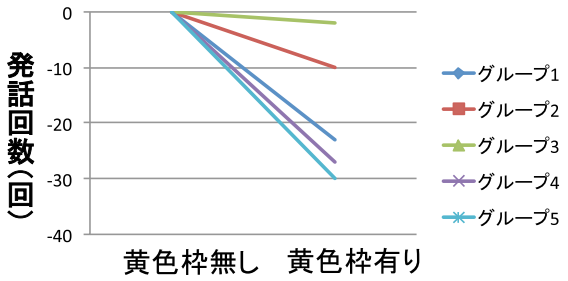

図 9 動作非教示群における，黄色枠有無による発話回数の比較：すべてのグループで，黄色枠を表示することにより，発話回数が減少した

Fig. 9 Number of utterances in unwitting groups.

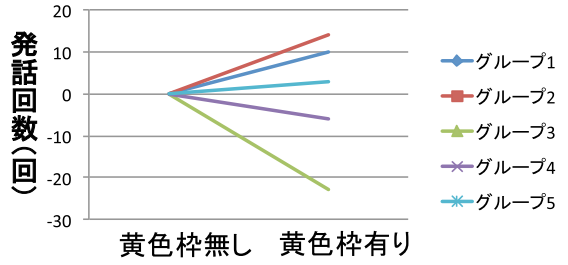

図 10 動作教示群における，黄色枠有無による発話回数の比較：黄色枠を表示することによる発話回数の増減の傾向は見られなかった

Fig. 10 Number of utterances in witting groups.

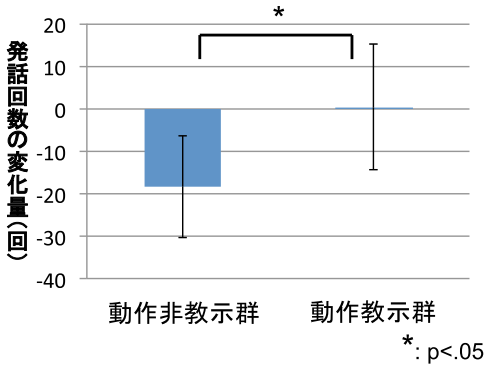

図 11 黄色枠有無による発話回数の変化量の比較：システムに検知される動作を教示しなかった群では，黄色枠を表示することにより発話回数が減少する傾向が見られた

Fig. 11 Comparison of shift of number of utterances.

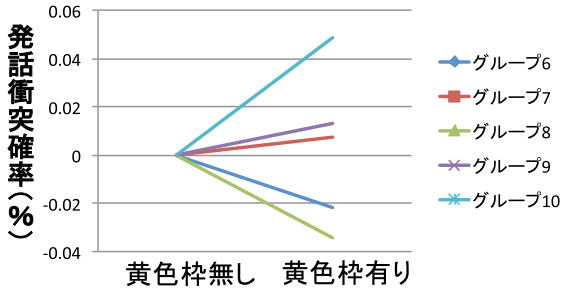

図 12 動作非教示群における，黄色枠有無による発話衝突確率の比較：黄色枠を表示することによる発話衝突確率の増減は見られなかった

Fig. 12 Speech contention rate in unwitting groups.

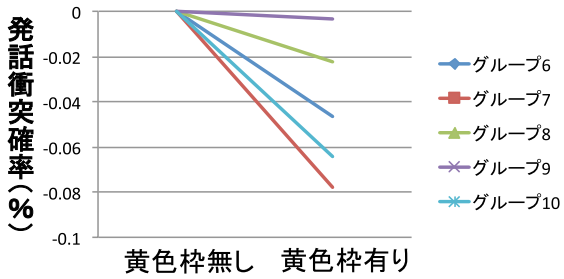

図 13 動作教示群における，黄色枠有無による発話衝突確率の比較：すべてのグループで，黄色枠を表示することにより発話衝突確率が減少した

Fig. 13 Speech contention rate in witting groups.

増加，減少の傾向は見られなかった．動作教示群では，5つのグループすべてが，黄色枠表示条件で，発話衝突確率が減少していた．動作教示群と，動作非教示群の間で，発話衝突確率の黄色枠なし/あり条件間の変化量の平均値を算出し図 14 にまとめた．これはすなわち，動作非教示群，動作教示群のうちどちらが，黄色枠を表示したときに，より発話衝突確率を低減させることができたかを比較した図である．両平均値の比較を t 検定で実施したところ，有意差 5%で，動作教示群の方が，黄色枠を表示したときに話衝突確率が減少することが分かった．

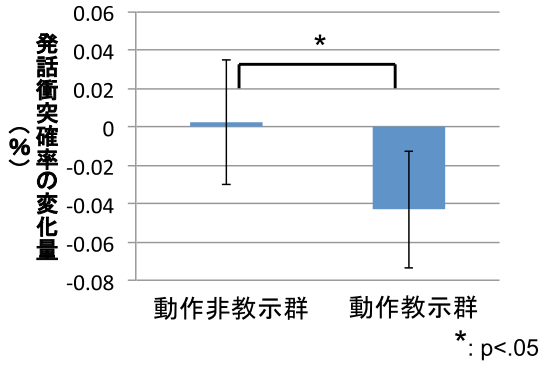

図 14 黄色枠有無による発話衝突確率の変化量の比較：システムが検知する動作を教示した群では，黄色枠を表示することにより発話衝突確率が減少する傾向が見られた

Fig. 14 Comparison of shift of speech contention rates.

7. 考察

本実験の目的は，次話者候補を強調することにより，発話が衝突する確率を低減させるという，本提案コンセプトの有効性を確かめることであった．「次に最も発話しそうな参加者に黄色枠が表示される．その判定のために，頷き，手の動き，頭の動き，相槌が検知されている」と教示した「動作教示群」においては，発話衝突確率が減少することが確かめられた．しかし，「次に最も発話しそうな参加者に黄色枠が表示される」とだけ教示された「動作非教示群」において，発話衝突確率は減少しないという結果が得られた．以下にこの原因を考察する．

図 15 のように，発話欲求が生じてから，発話にいたる流れを考える．具体的には，ある参加者に発話したいという欲求が生じ，その参加者が予備動作を行う．そしてその予備動作をシステムが検知する．そこから次話者候補として選定された場合，その参加者の映像に黄色枠が表示され，その参加者が判断をし，最終的に発話にいたる（もしくは，発話しない）．この流れに沿って発話するとき，発話欲求ステップで生じた発話欲求を，発話ステップへと 100%反映することができれば理想的である．しかし現実には，それぞれのステップ間の矢印区間になんらかのノイズが挟まれ，右のステップへ進むほど，発話欲求を反映できる確率が減少していく．発話欲求ステップと予備動作ステップの間には，参加者の性質やそのときの状況が影響し，予備動作を行うか否かが異なる．予備動作ステップと予備動作検知ステップの間には，予備動作の検知精度が影響する．予備動作検知ステップとシグナルステップの間には，次話者候補の選定精度が影響する．シグナルステップと発話ステップの間には，参加者がシグナル（黄色枠）を信用し，黄色枠が表示された参加者が実際に発話するか否か（黄色枠表示後発話確率）が影響する．

今回，発話欲求は測定できていない．そのため，参加者が予備動作である可能性のある動作を行った際に，それが

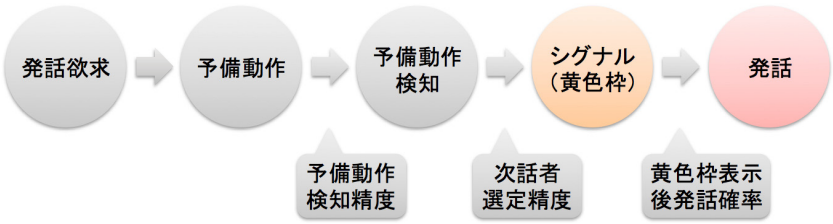

図 15 発話欲求が生じてから発話するまでのステップ
Fig. 15 Steps from generating the desire of speaking to speaking.

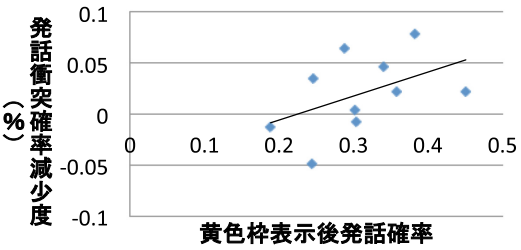

図 16 黄色枠表示後発話確率と発話衝突確率減少度の関係
Fig. 16 Relation between rate of speaking and decreasing level of speech contention rate.

予備動作であるか否かを測定することはできなかった．したがって，本プロトタイプシステムにおける予備動作の検知精度と次話者候補の選定精度を正確に測定する手段はない．本実験とは別に，予備動作である可能性のある動作の検知精度を求めた．2 人の被験者が，60 回ずつ，各予備動作を行い，システムに正しく検知された確率を求めるという，基礎的な実験を行ったところ，各々，手は 78.3%，頭は 65.8%，頷きは 64.2%，音声は 98.3%であった．

また，黄色枠表示後発話確率を，ビデオ解析データを基に，グループごとに算出した．黄色枠表示後発話確率と，黄色枠なし・あり条件間での発話衝突確率の減少度（発話衝突確率減少度）の関係を散布図に表した（**図 16**）．この 2 つの値の相関を求めたところ，相関係数は 0.47 で，中程度の相関があることが分かった．このことから，（黄色枠を使って，必ずしも発話できているわけではないが，）黄色枠を使って発話する（黄色枠が表示された直後に発話する）割合が高いほど，発話衝突確率は減少する傾向があることが分かった．

黄色枠表示後発話確率の，動作教示群 5 グループでの平均は 33.4%で，動作非教示群 5 グループでの平均は 28.6%であった．このことから，動作教示群の方が，黄色枠を使って発話する割合が高かったことが分かる．いい換えれば，動作教示群の方が，より黄色枠というシグナルを信用して発話したと考えられる．動作教示群の方が，動作非教示群よりも，黄色枠条件において発話回数が増えていることからも，動作教示群では，1) 話者が，黄色枠の表示されている他の参加者へ発話権を譲ったこと，2) 黄色枠の表示された参加者が躊躇せず発話したこと，などが考えられる．動作教示群において参加者が黄色枠を信用することができた理由としては，具体的な予備動作を教示されることで，積極的にその予備動作を行い，図 15 の発話欲求ステップと予備動作ステップ間のノイズが減少し，結果的に，黄色枠というシグナルが発話欲求を反映できる確率が高まったことが推測される．

今後は予備動作検知精度，次話者候補選定精度を上昇させていく．図 15 の流れにおいて，発話欲求ステップからシグナルステップまでの精度が高まるほど，参加者からのシグナルに対する信用が高まり，それを使用する頻度が高まり，発話衝突を低減させることができるという仮説が正しければ，十分に精度を高めることで，具体的な予備動作の教示・非教示にかかわらず，黄色枠を利用して発話権の授受を行い，発話衝突確率を低減できるようになり，より使いやすい会議システムを実現することが期待できる．

8. おわりに

本研究は，Web 会議のように映像の大きさ，解像度，またネットワークの速度に制限のある遠隔会議システムにおいて，発話の衝突確率を低減させることで円滑な話者交替を可能にすることを目的としている．特に発話の衝突が顕著に見られる，フリーディスカッションや，司会者の設定されない創造会議を対象としている．本稿では，Web 会議で発話前に現れる予備動作を抽出し，その活用度合いを調査した．そして，この予備動作を検知し，それを基に次の話者候補を選定し，全参加者へと提示することにより発話の衝突を低減させる手法を提案した．

プロトタイプシステムを構築し，評価実験を行った結果，システムが検知している具体的な予備動作を教示した動作教示群において，本手法を用いることで発話衝突確率が減少する結果が得られた．予備動作の検知精度，次話者候補の選定精度をより向上させれば，さらにユーザにとって自然な使用感と，発話衝突確率の低減を期待できる．上記 2 つの精度向上のため，機械学習アルゴリズムの導入を検討している．

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] 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).
[3] 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).
[4] Sacks, H., Schegloff, A.E. and Jefferson, G.: A Simplest Systematics for the Organization of Turn-Taking for Conversation, *Language*, Vol.50, No.4, Pt 1, pp.696–735 (1974).
[5] 高橋 誠：会議の進め方，日本経済新聞出版社 (1987).
[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] 松尾 隆：コミュニケーションの心理学，ナカニシヤ出版 (1999).
[8] 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).
[9] Sellen, A.J.: SPEECH PATTERNS IN VIDEO-MEDIATED CONVERSATIONS, *CHI'92*, pp.49–59 (1992).
[10] 入手先 ⟨http://www.meetingplaza.com/index-j.html⟩（参照 2011-10）.
[11] Cassell, J. and Vilhjálmsson, H.: Fully Embodied Conversational Avatars: Making Communicative Behaviors Autonomous, *Autonomous Agents and Multi-Agent Systems*, Vol.2, pp.45–64 (1999).
[12] 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).
[13] 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).
[14] Kawashima, H., Nishikawa, T. and Matsuyama, T.: Visual Filler: Facilitating Smooth Turn-Taking in Video Conferencing with Transmission Delay, *CHI 2008 Extended Abstract*, pp.3585–3590 (2008).
[15] Kelley, J.F.: An iterative design methodology for user-friendly natural language office information applications, *ACM Trans. Office Information Systems*, pp.26–41 (1984).
[16] 榎本美香：日本語における聞き手の話者移行適格場の認知メカニズム，ひつじ書房 (2009).

玉木 秀和

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．CSCW，ヒューマンインタフェースの研究に従事．現在，NTT サイバーソリューション研究所勤務，慶應義塾大学博士課程在学中．

東野 豪

1988 年東京工業大学大学院総合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 サイバーソ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，ヒューマンインタフェース，障がい者メディアの研究に従事．

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．CSCW，ヒューマンインタフェースの研究に従事．現在，NTT サイバーソリューション研究所主幹研究員．博士（工学），ACM，IEEE Computer Society，日本バーチャルリアリティ学会等の会員．

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 サイバーソ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．

岡田 謙一 （フェロー）

慶應義塾大学理工学部情報工学科主任教授，工学博士．専門は，CSCW，グループウェア，CHI．学会誌編集主査，論文誌編集主査，GN 研究会主査，日本 VR 学会理事等を歴任．現在，情報処理学会理事，電子情報通信学会 HB/KB 幹事長．情報処理学会論文賞（1996，2001，2008 年），情報処理学会 40 周年記念論文賞，IEEE SAINT'04，ICAT'07 最優秀論文賞等を受賞．情報処理学会フェロー，IEEE，ACM，電子情報通信学会，人工知能学会各会員．