Refereed Conference paper

人のコミュニケーションリズムに着目した映像通信メディア

玉木秀和[†]　中茂睦裕[†]　東野豪[†]　小林稔[†]
鈴木由里子[††]

世界不況の進行や環境負荷制約の高まりなどの社会情勢から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，低コストで導入でき，自席で会議に参加することができる手軽さがあり，市場も成長傾向にある．しかしその反面，Web会議システムでは会議参加者の映像が小さく分割された領域に表示されることで相手の様子を掴みにくく，ネットワーク遅延の影響で発話のタイミングを掴みにくいなどの要因により，発話の衝突が起こりやすい．これは，人は対面では頷きや相槌，視線，表情，身体動作などを使ってコミュニケーションを円滑化しているが，上述のような制約のあるWeb会議の環境ではこれを行うことが難しいからであると我々は考えた．本研究では，話者交替の場面に着目し，発話権を得ようとする動作をしている会議参加者を検知し，それを知らせることにより，スムーズに話者交替が行える環境の構築を目指す．本稿では，この研究の第一歩として本アプローチの妥当性を検証した．

Video Communication Media Focused on “Communication Rhythm”

Hidekazu Tamaki[†]　Mutsuhiro Nakashige[†]
Suguru Higashino[†]　Minoru Kobayashi[†]
and Yuriko Suzuki[††]

In Web conference, we sometimes hesitate when to start talking. Some people start talking at a time and then some of them stop talking. We think that, miscommunication of non-verbal messages and nod (using a voice) from small picture, low frame rate and network delay causes this problem. In this research, our aim is create an environment where we can change a speaker smoothly. To realize it, we propose “Video Communication Media Focused on “Communication Rhythm””. Our proposal system senses actions to have a right to speak and visualize it to other participants. In this paper, we verified a validity of our approach through experiment for a first step of this research.

1.　はじめに

世界的な不況の進行，感染症の流行，CO2排出削減意識の高まりなどの背景から，遠隔会議の需要が高まり，市場は成長傾向にある[1]．その中でも特にWeb会議は導入が手軽で,会議室に移動せず自席で実施できるため空間的制約も少なく,便利である．

しかしその反面Web会議には，自席のブラウザ上という環境で行うための制約もある．それはすなわち，通信帯域の保証されていないネットワークを通じて行うための遅延，フレームレートの低下や，デスクトップを分割して複数参加者の映像を表示するために一人当たりの映像が小さくなることなどである．Web会議では，映像よりも音声が優先されることが多いため，映像の質が悪くなりがちである．これらのことが原因で参加者の表情が読み取れない[2]というような問題が生じる．映像を使う利点の1つは非言語情報を伝えられること[3]であるが,その利点を活かすことができていない．

Web会議を行っていると，間が掴み辛く，発言するタイミングを計り辛い．そしていざ発言すると他の参加者と発言が衝突するという問題が起こる．この発言の衝突が起こらないようにしようとすると，沈黙時間が増えて白けた会議になってしまう．またWeb会議中に込み入った話になると，「会って話そう」という場面はよく見受けられる．人は対面した場面では非言語情報をうまく伝達しあうことでコミュニケーションを円滑に行っているが[4]，上記の理由によりWeb会議では映像チャネルを通じて充分な非言語情報の伝達ができていないことが，このような問題が起こることの大きな一因であると我々は考えた．

Web会議において発話の衝突が起こらずにスムーズに話者を交替しながら会議が進められるようになれば，衝突を避けるために沈黙し，白けることもなくなる．Web会議で活発に議論できるようになれば,Web会議自体の利用機会拡大が望め，冒頭に述べたような社会要請にも対応していくことができる．

このために本稿では，さまざまな制約のあるブラウザ上で行うWeb会議において，非言語情報によるコミュニケーションの円滑化を，特に話者交替の場面に着目しサポートする方法を提案する．本研究の学術的意義は，これまで音声と映像という2つのモダリティによってコミュニケーションを行っていたWeb会議に，「コミュニケーションを円滑化するための非言語情報のやり取り」というモダリティを追加する点にある．

† 日本電信電話株式会社 サイバーソリューション研究所
NTT Cyber Solutions Laboratories
†† NTT コムウェア
NTT Comware Corporation

2. 遠隔会議システムにおける非言語コミュニケーション

遠隔会議システムには，Web会議[5][6]のような小規模システムとテレビ会議，テレプレゼンスのような大規模システムがある．大規模システムであるテレビ会議システムやテレプレゼンスでは，高臨場な環境を追求し，非言語情報の伝達をある程度実現してきている[7][8]．大型ディスプレイを用い，解像度を高め，表示される人を等身大に見せ，視線が一致するような工夫を凝らしている．ディスプレイに映った自分以外の会議の参加者が，あたかも目の前にいるような存在感を得られる．しかし，会議の参加者の映像を等身大に映すことも,複数いる参加者の視線を一致させることも,Web会議環境で実現することは難しい.それぞれの参加者が自席で使用することができ,ブラウザ上で実行するWeb会議のシステムにとって,新しく大きなディスプレイを複数台用意することや，センシングデバイスを追加することは，誰もが手軽に使用できるというメリットを低減してしまう．

デスクトップ上で行うことのできる小規模なシステムの中には，デスクトップ上に仮想的な会議室を作り,会議参加者がアバタに扮して会議(会話)を行うものがある．これらはWeb会議に元々使われている設備を用いてシステムを構築することができる．渡辺らは，発話音声のON/OFFのリズムに基づいて頷きのタイミングを予測し，聞き手役アバタに頷き動作をさせている[9]．また藤田らは，発言の終わりに，他の参加者のアバタを注視する動作を擬似的に作り出し，話者交替を促す試みをしている[10]．しかしこれらのシステムでは，遠隔地にいる参加者の意図や反応に関わらず，擬似的に動作が作り出している．このため参加者の意図に反した頷きや注視動作が発生する可能性があり，それが認識のずれを起こし，コミュニケーションの失敗の原因になることが考えられる．

そこで我々は，Web会議を行うような小規模な環境（各参加者が自席にいながら，ブラウザ上で動作する環境）でも，遠隔地にいる参加者の非言語的なメッセージを効果的に伝え，話者交替をスムーズに行える方法を検討する必要があると考えた．

3. アプローチ

3.1 発話権を得ようとする動作

人は対面コミュニケーションで発話交代を行うとき，発話権を譲渡する動作や，発話権を得ようとする動作をする[4]．この内，発話権を得ようとする動作には頷き，胴体の動き，視線，相槌がある．具体的な例としてマジョリーは以下のような例を挙げている．「自分の方へ発言の機会をゆずってもらうためには，まず組んでいた足を元に戻し，腕組みもほどいて，身を前に乗り出し，次に話し手と対面するように，やや向きを変え，相手の視線を捉えようとする．そして，相手の話に応じて，ややめだつように頷いたり，同時に相手の発言内容への賛否を示すため「フン，フン」とか「ウーン」という声をやや大きくする[4]．」

上記の例における，「フン，フン」や「ウーン」という声に出した相槌は，音声チャネルさえ充分に確保されていれば他の参加者に認知される．この場合でも，全ての参加者の音声チャネルが多重されていることが条件であり，一度に発言できる参加者が限られているシステムや,ボタンを押さなくては発言できないシステムは除かれる．全ての参加者の発する音声が共有されている状況でも，音声を伴う相槌は，人が発声している様子がはっきり分かる映像がなければ，参加者人数が増えれば増えるほど，誰の相槌であるか分かり辛い．

また，上記の例における音声による相槌以外の「組んでいた足を元に戻す」，「腕組みをほどく」，「身を前に乗り出す」，「話し手と対面するようにやや向きを変える」，「相手の話に応じてややめだつように頷く」といった動作は，Web会議のような制約のある環境の映像では，認知することが困難である．我々はこれらのことが，Web会議において話者交替が上手くいかない問題の原因になっていると考えた．そこで，上に示したような発話権を得ようとする動作をシステムが検知し，それを他の参加者へ伝える手法を提案する．

3.2 人のコミュニケーションリズムに着目した映像通信メディア

人は対面コミュニケーションでは文章としてはっきり意味を持たない相槌や非言語情報を活用して会話を円滑化しているが[11][12]，それらは時系列的な関係性を持っている[13]．すなわち，単に非言語情報などを独立のタイミングで発するのではなく，コミュニケーションをとる相手の言動のタイミングに影響を受け，発するタイミングを選んでいる．例を挙げると，ある発話に対する頷き動作は，発話の呼気段落の終わりに出現することによってその発話を促し[13]，そのタイミングが前後することによって相手に受け取られる意味合いが変わるといったことである．

前節までに述べたように，Web会議を行う環境では，非言語情報を余すことなく伝達し合うことはできない．そこで，本提案手法では，発話権を得ようとする動作を検知して他の参加者へ伝えるのだが，ここで上記の点に着目する．具体的には，発話に対してタイミングの合っている「発話権を得ようとする動作」を検知し，誰がその動作をしているかを他の参加者へ伝える．発話に対してタイミングを合わせて頷いている動作や相槌，発話の後半に大きくなる胴体の動きを検知し，それらの動作が顕著に見られる参加者を他の参加者へ示す．まさに頷いているその様子をそのまま伝えることや，胴体の動きをフレームを落とさず伝えることは困難であるが，誰が発話権を得ようとしているかを伝えるだけであればやり取りする情報量は大幅に削減される．例えフレームレートが低くても，この小さな情報だけ素早く伝達してしまえばよい．本提案システムの構成のイメージを図1に示す．

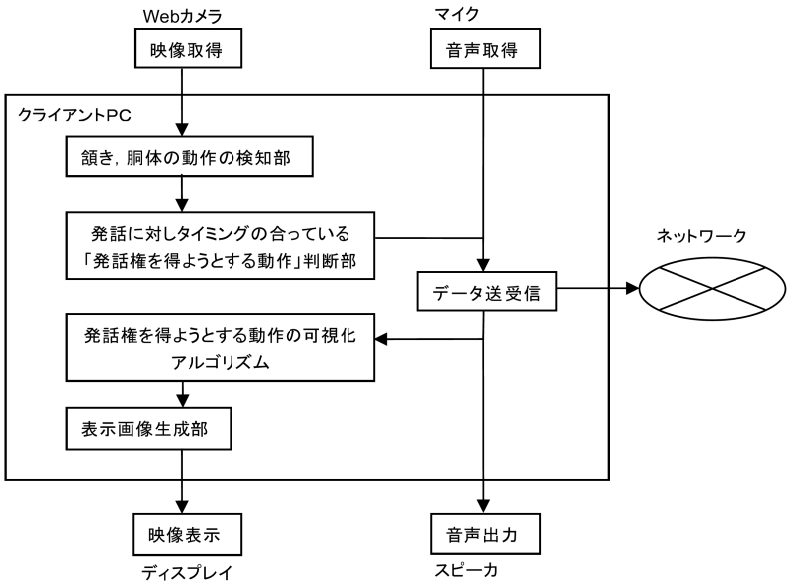

図 1　システム構成のイメージ

4. 予備実験

4.1 実験目的

本提案手法は，Web会議において，発話権を得ようとする動作をしている参加者を他の参加者へ知らせることで，話者交替がスムーズに行われ，活気のある会議を可能にすることが目的である．今回はまず予備実験という位置づけで，遠隔地にいる各参加者の映像が，ディスプレイ上の分割された区分にそれぞれ表示される環境で会議を行い，発話権を得ようとしている動作を見たときに，話者交替がスムーズに行われるかどうかを調べることを目的とした．特に発話にタイミングを合わせて頷いている，もしくは胴体の動作の大きい参加者1人の映像を強調して視線を向けさせた．今回は，発話権を得ようとする動作そのものを見せたときの効果を測るため，遅延が少なく，フレームレートも高い環境で行った．

4.2 実験環境

被験者3人と実験者1人からなる4人の参加者にそれぞれ個室のデスクトップ会議環境で会議をさせた．各個室には机があり，その上にマイク，カメラ，スピーカ，ディスプレイを1つずつ設置した.各個室のマイクとカメラで取得した映像は実験者の部屋へ集約し，実験者が操作して各個室へ配信できるようにした．参加者の音声をマイクから取得し，全参加者の音声を多重してスピーカから流した．また，参加者の胸から上の映像をカメラから取得し，実験者のいる部屋に設置されたPCに送り，そこで「田」の字に分割された画面上にそれぞれ表示した（図2）．1人あたりの映像の大きさはQVGAであった．この表示映像はVGA分配機を通し，全ての参加者のいる部屋に設置されたディスプレイへと映し出した．実験者の所持しているキーボードを操作することで，全ての参加者映像をカラー表示するモードと，特定の1人の参加者映像をカラー表示し，その他の参加者映像を白黒表示するモードを切り替えられるようにした．後者のモードでは，カラー表示する参加者映像をキー操作で選択できるようにした．

図 2　実験時ディスプレイ映像

4.3　手順

上記 4 人の参加者（被験者 3 人，実験者 1 人）で，6 分間の会議を行った．9 人の被験者の中で，3 人の被験者グループを作り，合計 3 回の会議を行った．会議の議題は被験者の身近な題材を選定し，「所内の食堂の改善案を考える」「独身寮の改善案を考える」といった内容であった．

6 分間の会議を 3 分間ずつ 2 つのパートに分け，それぞれのパートで参加者映像の表示方法を変えた．2 つの表示方法とはすなわち，通常の Web 会議のように全ての参加者の映像が平等に表示される方法（表示法 1）と，参加者の内，1 人の映像が強調して表示される方法（表示法 2）である（図 3）．1 人の像が強調して表示される表示法 2 では，強調される参加者は，特に発話にタイミングを合わせて頷いている，もしくは胴体の動作の大きい参加者とした．被験者にこのルールは知らせなかった．ある参加者の映像が強調されたら，その効果は次に特に発話にタイミングを合わせて頷いている，もしくは胴体の動作の大きい別の参加者が現れるまで持続させた．この強調表示の切替は実験者が行った．2 つの表示法は，3 回の会議の中で，順番を入れ替えて行った．

会議中のディスプレイの映像をビデオカメラで録画し，実験後に解析した．

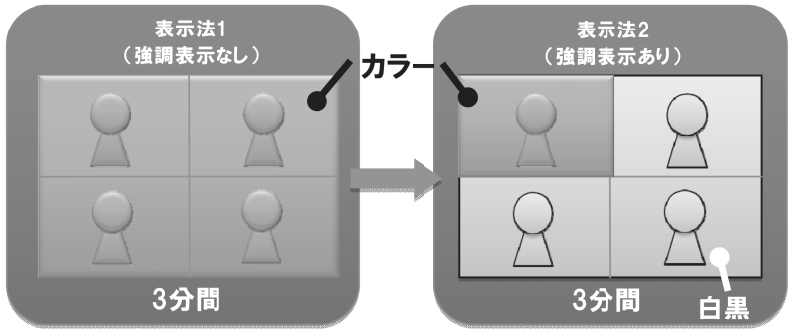

図 3　各表示法と強調表示

表 1　各表示法とカラー，白黒表示のルール

	強調表示	カラー/白黒表示
表示法 1	なし	全ての参加者映像をカラー表示
表示法 2	あり	特に発話にタイミングを合わせて頷いている，もしくは胴体の動作の大きい参加者営巣をカラー表示．その他の参加者映像は白黒表示

4.4　結果と考察

3 回の会議のうち，発言に偏りのなかった 1 つの会議をビデオ解析をした結果を表 2 に示す．発話衝突回数とは，同時に複数の参加者が発話し始め，そのうちいずれかの参加者が発話を中止した回数である．同時に発話し続けた場合は含まない．表示法 1，2 共に沈黙時間が短く，発話衝突回数が 1 回と少なかった．この結果から，会話は滞りなく行われ，盛り上がっていたことが分かる．映像は QVGA のサイズがあれば表情や身体の動きを認識することができ，充分なフレームレート（20〜30fps）と少ない遅

延であればその変化を読み取って，話し始めるタイミングを掴むことができたといえる．このことに関しては，実際の Web 会議環境では映像サイズ，フレームレート，遅延共に今回の環境よりも条件が下がることが普通であるため，次に行う実験では，実際のネットワーク環境にて上記結果がどうなるかを調べる必要がある．

次に，話者交替の回数は表示法 1 では 26 回，表示法 2 では 25 回と同程度であったが，強調表示後発話回数は表示法 1 では 8 回，表示法 2 では 13 回となった．強調表示後発話回数とは，カラー表示された参加者が次の発話権を得た回数である．表示法 1 では強調表示は行わなかったが，比較のため，表示法 2 と同じルールで強調表示させるべき参加者を選択し，その参加者が次の発話権を得た回数をカウントした．話者交替した回数のうち強調表示後に発話した回数の割合を，強調表示後発話回数欄の下部に記した．4 人で会議を行っている場面で，1 人の発話が終わって話者交替をする場合，次に発話する可能性があるのは 3 人である．すなわち，ランダムで話者交替をした場合には，1 人当たりの次に発話する確率は 33.3%となる．表示法 1 ではこの割合は 31%であった．これは，強調表示しなかった場合に，次に発話権が移る確率は，頷き動作や胴体の動きに関係ないことを示しているといえる．一方表示法 2 では，この割合は 50.2%と，表示法 1 の 1.4 倍となった．この原因は次の 2 つの可能性が考えられる．1 つ目は，1 人の参加者映像が強調表示されたため，その参加者が次に発話すべきだというある種の強制力を，自他共に感じたという可能性である．これは会議をする場面で良いことか悪いことかは，現状では判断できない．2 つ目は，ある参加者が他の参加者の発話に対してタイミングよく頷き，胴体の動きが大きくなった様子が強調され，そこに参加者たちの注目が集まり，発話権が移譲されたという可能性である．これは人が対面コミュニケーションで行っている話者交替の自然な流れに近づけたことを示している．こちらが正しければ，発話に対して特にタイミングを合わせて頷いている，もしくは胴体の動作が大きい参加者の映像を強調することで，会議における話者交替をスムーズに行うことができるといえる．

被験者のコメントから，特に頷いている参加者，よく聞いている参加者が分かったということが得られた．また，次に話そうと思ったときに自分が強調表示される場面がよくあった，強調表示されると注目されている感じがする，というコメントも得られた．

表 2　ビデオ解析結果

	発話時間	沈黙時間	発話衝突回数	話者交替回数	強調表示後発話回数
表示法 1	3 分 00 秒	0 秒	1 回	26 回	(8 回) (31%)
表示法 2	2 分 57 秒	3 秒	1 回	25 回	13 回 (50.2%)

5. 今後の検討事項

今後の検討事項としては，追実験について，発話権を得ようとする動作の検知，そしてその可視化についてである．

5.1 追実験

今回行った予備実験では，ネットワークを介していないため，遅延が少なく，フレームレートも高かった．また，映像の大きさに関しても，QVGAと，Web会議にとっては大きいものを用いた．これだけの環境を整えて，特に発話にタイミングを合わせて頷いている，もしくは胴体の動作の大きい参加者の映像を強調して見せることで，スムーズに話者交替を行い活発に会話を行うことができそうだということが分かった．しかし実際のWeb会議の環境では，ネットワークを介すために遅延が生じ，フレームレートが低くなる．また，複数人（6～10人程度）で行い，資料を共有しながら会議を行うと，1人あたりの映像領域は小さく限られたものになる．そこで，次回の実験では，実際のWeb会議環境の制約がある状況で，特に発話にタイミングを合わせて頷く，もしくは胴体の動作が大きい参加者が誰であるかを示したときに，沈黙時間や発話衝突回数がどう変化するか，話者交替にどう影響を及ぼすかを調べる必要がある．すなわち，今回は発話権を得ようとする動作そのものを伝達して見せていたが，それは実際のWeb会議の環境下では実現が難しい．そこで誰が発話権を得ようとする動作をしているのかを伝達することによる効果を測るための実験を行う．

5.2 発話権を得ようとする動作の検知

発話権を得ようとする動作そのものを伝達することはできないため，これをシステムが検知する仕組みを検討する．これには，頷く，胴体を動かすという動作自体の検知と，そのタイミングが発話に合っているか，生起すべきところで生起しているかを判別することが必要である．

5.3 発話権を得ようとする動作の可視化

発話権を得ようとする動作をしている参加者を示す方法を検討する．今回の予備実験では，1人の参加者映像のみをカラー表示し，それ以外を白黒表示することによって

これを行った．これ以外にも1人の参加者映像を強調する方法は多数考えられるが，どの方法が本提案手法に適しているかを検討する必要がある．強調方法により，誘目性などに差が出ることが考えられる．会議の進行を妨げずに可視化する手法を構築する．

6. おわりに

Web会議において，発話の衝突が起こらずにスムーズに会話が進み，会議が活性化される環境を目指して，発話にタイミングを合わせて頷いている，もしくは胴体の動作が大きい参加者の様子を，他の参加者へ強調して知らせるアプローチを提案した．

まず予備実験として1人あたりの映像サイズがQVGA，低遅延，高フレームレートの環境で，発話にタイミングを合わせて頷く，もしくは胴体の動作が大きい参加者映像を強調してその様子を見せることで，話者交替の場面での影響を調べた．実験結果から，この環境では発話に躊躇することや，発話が衝突するという状況は見られなかったが，発話にタイミングを合わせて頷く，もしくは胴体の動作が大きい参加者映像を強調することで，スムーズに話者交替ができる可能性があることが示された．

今後の検討事項として，まず実際のネットワーク環境で追実験を行い，本提案アプローチの妥当性を確かめるとともに，発話権を得ようとする動作を検知し，可視化する技術を検討，実現していく．

謝辞　本研究の実験のアドバイスを下さった皆様，被験者の皆様，そして論文の校閲をして頂いた皆様に，謹んで感謝の意を表する．

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 年 3 月
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) Ellen A. Isaacs, John C. Tang, "WHAT VIDEO CAN AND CAN'T DO FOR COLLABORATION: A CASE STUDY", ACM Multimedia 93, pp.199-206, 1993.
4) マジョリー・F・ヴォーガス：非言語コミュニケーション，新潮社，1987.
5) http://www.ntt.co.jp/journal/0507/files/jn200507013.pdf（2009 年 7 月現在）
6) http://www.meetingplaza.com/index-j.html（2009 年 7 月現在）
7) http://h20341.www2.hp.com/enterprise/cache/570007-0-0-109-200.html（2009 年 7 月現在）
8) http://www.cisco.com/web/JP/solution/telepresence/index.html（2009 年 7 月現在）
9) 渡辺富夫，大久保雅史，石井裕，中林慶一：バーチャルアクターとバーチャルウェーブを用いた身体的バーチャルコミュニケーションシステム，ヒューマンインタフェース学会論文誌, Vol.2, No.2, pp.1-10, 2000.
10) 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.
11) Ray L. Birdwhistell:”Kinesics and Context”, University of Pennsylvania Press, 1970.
12) 松尾隆：コミュニケーションの心理学，ナカニシヤ出版，1999.
13) 渡辺富夫，夏井武雄：ヒューマン・インタフェースへの音声対話時の引き込み現象の応用に関する研究：うなずき反応を視覚的に模擬する音声反応システムの開発，昭和 63 年度厚生省心身障害研究「家庭保険と小児の成長・発達に関する総合的研究」, pp.64-70