

## 話者交替時に使用される非言語情報は 会議環境により変化するか

玉木秀和<sup>†</sup> 東野豪<sup>†</sup> 小林稔<sup>†</sup> 井原雅行<sup>†</sup>

Web 会議システムを用いてブレインストーミングなどの創造会議を不自由なく実施可能にするために、発話の衝突を回避する仕組みの実現を目指している。本稿では、発話が衝突する原因を究明するために、さまざまな条件を設定した実験環境で会議を行い、話者交替時の非言語情報の出現傾向を調査した。その結果、頷きや相槌などの、意図的に発話権を獲得しようとする際に利用される非言語情報の出現頻度が下がることが分かった。今後はさらにこの原因を明らかにするための実験を続ける。

## Do They Use Different Set of Non-verbal Language in turn-taking in Distributed Conferences?

Hidekazu Tamaki<sup>†</sup> Suguru Higashino<sup>†</sup>  
Minoru Kobayashi<sup>†</sup> and Masayuki Ihara<sup>†</sup>

Our aim is create an environment where we can change a speaker smoothly. To avoid speech contention, we try to investigate what is the main reason of it. In this paper, we experimented to investigate what kind and how many non-verbal information appeared in distributed conferences. The result said that, non-verbal information such as nodding and back-channel feedback appeared decreasingly, even though they did conferences with video conference, nc delay.

### 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]。遠隔会議システムの中でも特にWeb会議は、低コストで導入でき、自席にしながら会議に参加することができる手軽さがある。しかしその反面、画像の大きさや解像度、ネットワークの伝送速度などに制約があり、さらにその環境は会議の参加者ごとに異なっている。そのため他の会議参加者の様子が分かりづらい[2]、映像や音声に遅延があるのでタイミングを計りづらい[3]など様々な要因により、発話の衝突が起こりやすい。Web会議システムを用いた会議では、周知会議や簡単な意見交換をすることに弊害はないが、深い議論や、ブレインストーミングのような創造会議をすることは難しく、「実際に会って議論しましょう」という場面はオフィスでよく見かける風景ではないだろうか。

人は普段、対面環境で会話を行う際に、非言語情報をうまく用いて衝突を回避している[4][5][6]が、Web会議だとこれができない。我々はこの原因を突き止め、Web会議においても、自然と話者交替時に発話権の授受ができるような仕組みの実現を目指している。そして、Web会議システムを用いて、遠隔地にいながらも手軽に創造会議を不自由なく実施できるようにすることで、Web会議システムの利用機会を拡大するとともに、冒頭に述べた社会要請に応じていく。我々はこれまでに、次に話そうとしている会議参加者の非言語情報（予備動作）をセンシングし、次に最も発話しそうな会議参加者を選定し、全参加者へ示すことで発話の衝突確率を低減させるアプローチをとり、一定の効果を確認した。

本稿では、発話の衝突が生じる要因を明らかにするためのステップを整理し、そのうちの第一ステップである「Web会議において予備動作はなくなるか否か」と、その原因を考察することを目的とする。そのために、会議環境を変化させることにより、話者交替時に使用される予備動作の使用頻度の変化を調査する実験を行い、分析した。

以降、発話の衝突要因を明らかにする本研究のステップを整理し、後半では今回実施した第一ステップの実験について報告する。

### 2. Web 会議における発話の衝突低減に向けたアプローチ

#### 2.1 課題の整理

今井らは、あるプロジェクトのメンバで会議を行う際に、会議の内容と選択されるコミュニケーションツールの関係性を調査した[7]。そして、複雑な会話構造の必要な会議ほど、遠隔会議よりも対面会議が好んで選択されることを報告している。

このように、一般的に Web 会議で実施が難しいと判断されるブレインストーミング

<sup>†</sup> 日本電信電話株式会社 サイバーソリューション研究所  
NIPPON TELEGRAPH AND TELEPHONE CORPORATION NTT Cyber Solutions Laboratories

のような創造会議は、短時間のうちに頻繁に話者交替が起こる会議である[8]。小磯らは、話者交替の失敗は、2人以上の会議参加者が同時に発話を開始する「発話の衝突」と、ある会議参加者の発話が終了した後、1.7秒以上無音状態の続く「沈黙」である、としている[9]。我々のこれまでの分析結果では、Web会議システムを用いてブレインストーミングを実施した場合、対面会議で実施した場合と比較して、話者交替時に発話が衝突する確率が20から30倍高まった[10]。このことから、我々はWeb会議でブレインストーミングを実施する際の大きな障壁の1つは、話者交替時に頻発する発話の衝突であると考えている。

人は普段の対面コミュニケーションでは発話権を渡す、獲得するという行為を、非言語情報を活用して行い、円滑に話者交替している。話者交替とは、発話権を受け渡す行為であるから、それに関わる、

- a) 発話権を渡すための非言語情報
- b) 発話権を獲得するための非言語情報

を絶妙にやり取りする必要がある。石井らはa)発話権を渡すための非言語情報である視線に着目し、これを擬人的に作り出すことで話者交替時の沈黙を減らす試みを行った[11]。また、マジョリーは、人は発話権を獲得する際に大きく傾くことや、目立つように相槌をうつことをする、と指摘している[4]。伝らも、話者交替の際に、次の発話権を獲得する参加者、すなわち「次話者」の非言語情報が大切であることを主張している[5][6]。a)b)どちらの非言語情報も重要な役割を担っているが、我々は本研究において、まずはb)発話権を獲得するための非言語情報を対象とし、Web会議における活用度合いを調査し、発話の衝突要因究明へ向けて研究を進める。

本研究において、b)のように、発話権獲得の意思がある、ないに関わらず、発話の前に現れる非言語情報を「予備動作」と呼んでいる。本稿では、この予備動作とWeb会議における発話の衝突との関係を究明するために、Web会議では予備動作は、

- (ア) 行われている
- (イ) 行われていない

のどちらであるかを調査する。この次の研究ステップとしては、もしWeb会議で予備動作が(ア)行われているなら、予備動作が他の会議参加者に認知されているかどうかを調査する。予備動作が他の会議参加者に認知されているならば、話者交替のプロトコルにおける予備動作が認知された後の段階で問題が起きている可能性と、そもそも予備動作が認知されるかどうかに関わらず、Web会議では発話が衝突してしまうという、予備動作以外の要因が関係している可能性とを検証する。予備動作が他の会議参加者に認知されていないならば、その原因を調査する。

もしWeb会議で予備動作が(イ)行われていないことが明らかになったならば、その原因と、どうしたら予備動作が行われるようになるか、に関する調査を行っていく。当然、予備動作を行ったとしても、他の会議参加者に認知されないから予備動作を行

わなくなる、というように要因が相互に関係し合うことも想定される。この場合、予備動作を行わなくなるのは、

- ・他の参加者の予備動作を認知することができないため
  - ・自分の予備動作が他の参加者に伝わっていないと感じるため
- のどちらが原因であるかを検証する必要がある。

### 3. 実験

#### 3.1 目的

前章で述べたように、Web会議で予備動作が行われているか、行われていないかを調査し、またその結果について考察する。そのために、会議の条件を複数設定し、条件の違いによる予備動作の出現頻度、活用度合いを調査した。

#### 3.2 方法

##### 3.2.1 会議の構成と議題

56人の被験者を集め、実験を行った。4人一組の合計13グループを作り、会議を行わせた。会議の内容は、高橋の分類による創造会議の1つ、ブレインストーミングである。議題を与え、一定時間の間、アイデアを発散させるプロセスを続けさせた。議題は、

- ・未来の車を考える
- ・未来の家を考える
- ・未来のテレビを考える
- ・未来の携帯電話を考える

とした。上記4つの中から、グループ内のメンバが多数決をし、最も支持された議題についてブレインストーミングを行わせた。

1セッション10分間とし、セッションごとに実験条件を変えた。

##### 3.2.2 実験条件

実験条件は以下の5条件を設定した。グループごとの条件の設定順序はランダムとした。

- 1) Face-to-face
- 2) Curtain
- 3) Monitor
- 4) Web Conf.(200msec delay)
- 5) Web Conf.(300msec delay)

それぞれについて以下に説明する。

##### 1) Face-to-face

普段我々の行う対面会議である。テーブルを囲んで4人の会議参加者(被験者)

が向かい合って会議を実施した。なお、各会議参加者の前にはビデオカメラが設置してあり、実験記録のために各々の胸像を撮影していた。

## 2) Curtain

1) の条件に加え、各会議参加者の間に十字型のカーテンを設置し、互いの様子が一切見えないようにした。なお、音声はリアルタイムに通じていた。

## 3) Monitor

2) の条件に加え、各会議参加者の前に、モニタを設置した。ビデオカメラにより撮影された各会議参加者の胸像を、田の字状に4分割したモニタ画面へ表示させた。

## 4) Web Conf.(200msec delay)

各会議参加者は個別の防音室に入り、NTT アイティ社製の Web 会議システム MeetingPlaza[12]を利用して会議に参加した。各会議参加者はヘッドセットを装着し、システムを介してのみ音声のやり取りができるようにした。各々の正面に配置されたモニタの上部に Web カメラを取り付け、胸像を撮影した。

MeetingPlaza の設定は以下の通りである。

- ・ 画像のフレームレート：30fps
- ・ 画像の解像度：640x480 (VGA)
- ・ 画像の配置：4分割 (田の字状)
- ・ 音声遅延：200msec

## 5) Web Conf.(300msec delay)

4) の条件の内、音声遅延のみを 300msec に設定した。

### 3.2.3 解析内容・方法

先に述べた条件の内、1) 2) 3) は、ビデオカメラにより撮影した各会議参加者の胸像を、田の字状に4分割したモニタ画面へ表示し、これを録画した。また、条件4) 5) については、MeetingPlaza 画面を記録した。それぞれ10分間のセッションのうち、最初と最後のいくらかの時間は、条件切り替えに影響され、発話内容が議題からそれる場合があった。そのため、発話内容が議題に沿っている、中間の7分間を対象に解析を行った。解析内容とその方法について以下に述べる。

#### ● 予備動作の出現回数

記録した映像を分析し、各会議参加者の予備動作にタグ付けた。このタグ付けは、2人以上のタグが一致したものを採用した。

タグ付けた予備動作は以下のものである。

- ・ 頷き
- ・ 相槌
- ・ 首を傾ける
- ・ 体(頭)を前に乗り出す

- ・ 体(頭)を後ろに動かす
- ・ 体(頭)を横に動かす
- ・ 手で何かを例示する(形や動作)
- ・ 手を口、鼻周辺へ持っていく
- ・ 手を頭部へ持っていく
- ・ 手を挙げる
- ・ 上の3つのいずれかから手を下ろす
- ・ 手を前に出す
- ・ 下を向く
- ・ 顔を上げる
- ・ 話者へ視線を向ける
- ・ 次の話者へ視線を向ける
- ・ 話者、次話者以外の者へ視線を向ける
- ・ 笑う

なお、分析対象は参加者ごとの発話頻度が特に均等であった4グループ、16人の被験者とした。

#### ● 予備動作ごとの発話確率

上記それぞれの子備動作を行った後に、発話をした回数を計測し、その割合を求めた。

予備動作ごとの発話確率 = 予備動作後に発話した回数 / 予備動作回数

#### ● 発話が予備動作を伴う確率

発話が予備動作を伴う確率 = 予備動作を伴った発話の回数 / 発話回数

#### ● 質問紙

被験者全員 (56人) に対して、以下の項目について質問紙を用いた5段階の官能評価を行った。

- ・ 参加者の表情が分からない
- ・ 参加者のしぐさが分からない
- ・ 話し始めようとする素振りが分からない
- ・ いつ話し始めていいか分からない
- ・ 話し始めが他の人とぶつかる
- ・ 話を聞いているか分からない
- ・ 沈黙が多い
- ・ 話に割り込めない
- ・ 盛り上がらない
- ・ 視線を合わせられない
- ・ 緊張する

- ・ 相手の存在感が薄い

これらの項目が当てはまると思う場合は 5, 当てはまらないと思う場合は 1 を回答させた。

#### 4. 結果

##### ● 予備動作の出現回数

4 つのグループが行った会議の 7 分間に出現した予備動作の合計回数をグラフにまとめた (図 1)。全体的に頷きや相槌の出現回数が多いが、頭や手を用いた予備動作に関しては、さらに詳細に分けてタグ付けしているために、個々の値は小さくなっている。なお、Face-to-face 条件以外では今回は視線の動きを検知することはできなかった。

被験者のグループ単位で予備動作出現回数が条件によりどう変化しているかをまとめた (図 2-7)。なお、使用する身体の部位ごとに予備動作をまとめて、出現回数を集計した。頷きと相槌は、どのグループも概ね Face-to-face, Monitor, Curtain, Web の順に、出現回数が減少する傾向が見られた。頭の動きによる予備動作、手の動きによる予備動作、そして笑いはその出現回数に、条件による傾向は見られなかった。

##### ● 予備動作ごとの発話確率

母数が少ないものがあるのではらつきはあるが、どの条件でも予備動作ごとの発話確率は 4 割程度であった。4 人で会話をした場合、1 人が話者であるときに、次話者になる可能性がある参加者は 3 人になる。そのため、ランダムに発話をした場合には、発話確率は 3 分の 1 となる。つまり、発話確率が 4 割でも、ランダムに発話した場合と比較すると、発話する確率が高まったということになる。

Face-to-face の条件よりも、2 つの Web 会議環境の方が概して予備動作ごとの発話確率が高くなった。

##### ● 発話が予備動作を伴う確率

Face-to-face と Monitor 条件に比べ、Curtain, Web(200delay), Web(300delay)の順に発話が予備動作を伴う確率は減少した (図 8)。

##### ● 質問紙

質問紙による官能評価の結果を図 9 にまとめる。ほとんどの項目について、Face-to-face, Monitor, Web, Curtain の順に、値が高まった。

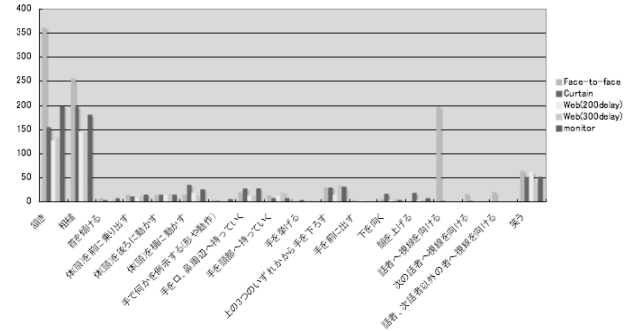


図 1. 予備動作回数

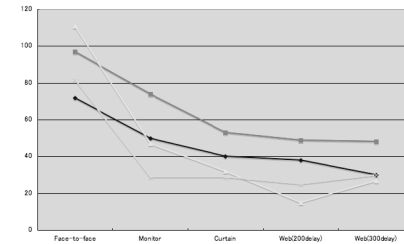


図 2. 頷き回数の比較

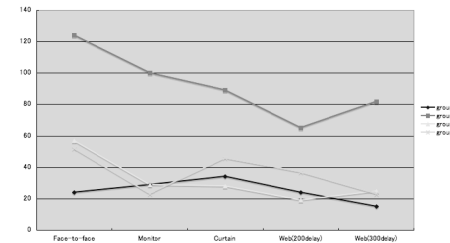


図 3. 相槌回数の比較



