2009－GN－71 (28)
2009／3／19

テレビ会議において視線の伝達が話者交替に及ぼす影響の分析

西村圭亮 上野晃嗣 坪井創吾 下郡信宏

株式会社東芝 研究開発センター

テレビ会議において視線を伝達することでスムーズな話者交替がどの程度実現できるのか，視線を伝達できるテレビ会議システムと視線を伝達できないテレビ会議システムを定量的に比較した．視線を伝達できないテレビ会議システムは，さらに高画質と低画質の画質の異なるテレビ会議システムを用いた．この結果，視線を伝達できないと意図しない相手が発言することが多くなった．またテレビ会議の画質の違いが，話者交替に影響を与えないことが分かった．以上から，テレビ会議において視線を伝達することで話者交替がスムーズに実現できることを確認した．

Analysis of Turn-Taking in Video Conference with Eye-Contact

Keisuke Nishimura, Kouji Ueno, Sougo Tuboi, and Nobuhiro Simogori

Corporate Research & Development Center, TOSHIBA Corporation.

Eye-contact is known to coordinate turn-taking in conversation. However, eye-contact is unable in a typical video conferencing system. An experiment to measure how eye-contact affects turn-taking in video conferencing is conducted. The experiment involved 52 subjects and three video conferencing systems, one with eye-contact and two without eye-contact which had different display resolution. Through the experiment, turn-taking errors increased significantly without eye-contact, while display resolution showed no significant difference.

1. はじめに

近年，企業のグローバル化，テレビの高画質化，ブローバンド普及に伴い，テレビ会議の需要が高まっている．テレビ会議のハードウェア性能が向上するなかで，カメラとディスプレイを1つずつ用いたテレビ会議の設定は変わっていない．一般に，1 人の出席者を撮影するカメラが 1 つだけの場合，多人数が参加するテレビ会議では，視線や顔向きの方向が実際に対面で行う会議との方向と一致しないために誰が誰を見ているのか，誰が誰に対して発言しているのか分からない．そのために伝えたい相手に話しが伝わらない，会話のタイミングが掴めないなどコミュニケーションがスムーズにいかない問題がある．これまでの研究から視線や顔向きは会話において重要な要素だと言われている[2][3][4]．従って，視線や顔向きの伝達が正しく行われるテレビ会議では，会話の流れ，つまり話者交替の流れがよりスムーズに行えるはずである．我々は，視線を伝達できるテレビ会議においてスムーズな話者交替が，どの程度実現できるか検証するため，視線が話者交替に及ぼす影響を定量的に評価する．

以下，2 章で視線に関する先行研究を述べ，3 章で実験の詳細を説明する．4章では，実験の分析結果について述べ，5 章でその結果を受けて考察する．

2. 関連研究

テレビ会議では視線が伝達できないことに関して以下のような問題点が指摘されている．複数の参加者と視線を合わせることができない．誰が誰に向かって注意を向けているのか分からない[5]．これらの問題を解決するために視線を正しく伝達するシステムの試作と評価が行われている．岡田らは，特殊なスクリーンを用いて視線を正しく伝えることができるテレビ会議システム”MAJIC”を開発し、被験者を用いて主観評価を行っている [11]．また Nguyen らは，再帰性反射材を用いて多人数同士でも正しく視線が伝わる会議システムを提案し，そのシステムを用いて視線がどの程度一致するかの評価を行っている[9]．しかし視線の一致が会話に及ぼす影響までは評価していない．新井らは，視線の伝達によって人間の補償行動がどのように表出するかを定性的に分析している[10]．

定量的な評価を行っている研究では，Sellen，

Vertegaal らの研究がある. Sellen は, 小型のディスプレイ, スピーカー, カメラが備わった Hydra Unit(以降, Hydra)と呼ばれる対面設定と同じ視線関係を再現するシステムを用い, 視線が正しく伝わらない会議システムとの比較実験を行った[5]. その結果, 同時発話や発言権のコントロールに有意差が見られたものの, 話者交替に差を発見することができなかった. これを受けて Vertegaal らは, 視線の受け渡しができない場合に話者交替数が 25%減少すると報告をしている[6][7]. しかしながら, これらの実験ではタスクの自由度が高く、被験者の行動の差が作業効率の差に必ずしも直結しない.

3. 実験

テレビ会議において, 視線が伝達できる場合と視線が伝達できない場合において, 話者交替の行われ方に違いが出ることを定量的に測定するために以下の実験を行った.

実験では, 定量的に違いを測定するため, 強制的に話者交替が発生するようにコントロールされたタスクを被験者に与え, タスクに掛かった時間と話者交替に失敗した回数を測定した.

被験者は, 視線が伝達できるテレビ会議1種類と視線が伝達できないテレビ会議二種類と対面を含めた4つの会議設定で上記のタスクを行ない, 会議設定の違いを分析した.

以下, 3. 1節では, 視線が伝達できる場合, できない場合に考えられる仮説を述べ, 次に 3. 2節にタスクや実験設定などの実験方法を説明する. 最後に 3. 3節で被験者に実施したアンケートについて述べる.

3.1 仮説

視線が伝達できた方が, スムーズに会話が行えることを示すため, 以下の3つの仮説を立てた.

仮説 1

視線を伝達できないと, 意図しない相手が発言することが多くなる

視線を伝達できない場合, 視線を伝達できる場合に比べて意図しない相手の発言が有意に増えることを確認する.

仮説 2

視線を伝達できないと, タスクに掛かる時間が増える

視線を伝達できない場合, 視線を伝達できる場合に比べてタスクに掛かる時間が有意に増えることを確認する.

仮説 3

画質の違いは, 話者交替に影響を与えない

画質の違いが話者交替に有意な差がないことを確認する.

3.2 実験方法

実験はグループで行い, 与えられた数字を小さい順に並べるタスクを4つの模擬 TV 会議システムで行った. 特定のテーマに関して自由に会話するのではなく、発言内容を数字に限定することで内容の複雑さの影響を排除することを狙った.

被験者

被験者は, 面識がある社内の研究者 52 名である. 実験は, 4 人で1グループとして行ない, 全部で 13 グループが実験に参加した.

タスク

被験者には1から 999 までの 12 個の数字が書かれた数字カードが渡される(図1). グループの 4 人の被験者には異なる数字が書かれた数字カードが渡される. 被験者は互いの数字カードの内容を知らない. グループに渡された合計 48 個の数字を小さい数字から順に間違えずに短時間で読み上げることを目標とする. 順番を間違えて, 数字を飛ばして読み上げてしまった場合(以降, 発言ミス)には, 飛ばされた数字を読み上げたのちに再度、間違えて読み上げてしまった数字を読み上げ直す. 数字カードには数字以外に次の数字を持っている人の座席の位置を示す

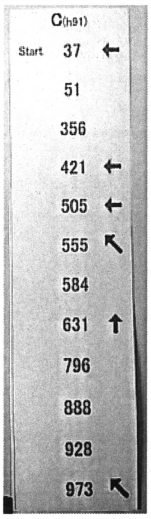

図 1. 数字カード

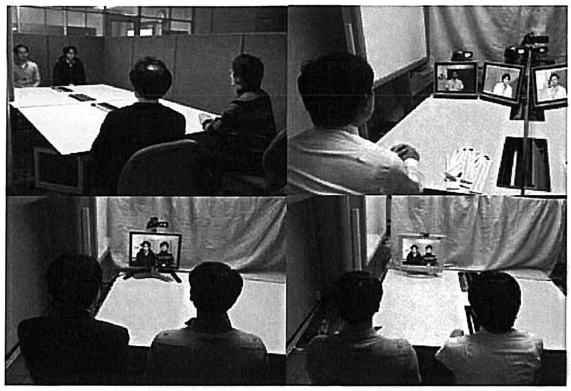

図 2. FTF(左上), Hydra(右上)、TV（左下）, Web(右下)

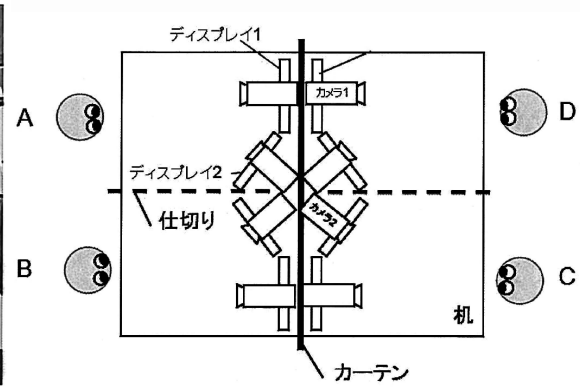

図 3. Hydra 設定

矢印が書かれているものがある．矢印が書かれていた場合には，その座席に座っている被験者に対して視線で発言を促す．読み上げの際に数字以外の言葉を発してはならず，手振りや身振りなど通常の会議では行わない行動を取ってはならないとした．

48 個の数字の読み上げを 1 セッションとして，会議システムや席順を変えて合計 12 セッション行った．1 つのセッションには平均 1 分 58 秒を要した．

模擬テレビ会議システム

実験は3種類の模擬的なテレビ会議システムを用いて行った．2つの地点で会議を行っていることを想定して被験者は2対2に分かれている．基準とするためにテレビ会議システムを使わない対面での実験も行った．以下に対面を含めた4種類の会議設定について説明する．なお，いずれの設定においても4人の被験者はカーテンで仕切られた同一の部屋で実験しており，音声はシステムを介さずに互いの発言が直接聞こえる．模擬テレビ会議システムは反対側にいる二人の被験者の映像のみを伝える．これにより，音声の品質，遅延，位置情報などの影響を排除することを狙った．

対面設定(FTF)

対面設定ではディスプレイを介さず 2 人 1 組で机の両端に座って対面する(図 2　左上).

Hydra設定(Hydra)

Sellen の研究の Hydra[5]を模したシステムである．各被験者の前に 2 枚のディスプレイがあり，対面に座っている被験者を 1 人ずつ写している．合計 8 枚のディスプレイを使って視線の伝達を可能にしている(図 2 右上). Hydra の設定を図 3 に示す．被験者 A の前には 2 つのディスプレイ 1 とディスプレイ 2 がある．

ディスプレイ 1 はカメラ 1 で正面に座る被験者 D を正面から写している．ディスプレイ 2 はカメラ 2 で斜め前に座る被験者 C を斜めから写している．これにより，被験者 A は被験者 C または D が自分の方向を見た時には見られていると分かる映像が表示される．逆に被験者 C または D が被験者 B の方向を見ている時には，被験者 A には自分は見られていないとわかる映像が表示されている．これにより視線の伝達が行われる．

テレビ会議設定(TV)

一般的な SD 画質のテレビ会議システムを模したシステムである．(図 2 左下). カメラとディスプレイを 2 組使い互いの映像を伝える．画質は 720x480 ピクセル 60fps．一枚のディスプレイに対面に座る二人を写している．ディスプレイに映る相手が誰を見ているのか分かりにくく，視線を伝達できないシステムである．

Web会議設定(Web)

テレビ会議設定と同じだが，Web 会議で使用される Web カメラを用いて低画質である (図 2 右下). 画質は QVGA(320x240 ピクセル, 30fps)である．

実験手順

1 つの模擬会議システムで 3 セッション行う．これを1セットと呼ぶ．

実験開始前に対面環境で被験者に実験方法の説明を行う．質問があれば回答する．続いてFTF で練習セッションを行う．被験者の希望があれば、3 度まで練習セッションを行う．今回の実験では 2 度の練習まで希望があった．練習セッションに引き続き、FTF での実験を1セット行う．セッション毎に被験者の席順を入れ替えてセット内では毎回違う被験者が隣に座るようにした．FTF での

実験が終わると模擬会議システムを用いた実験を行う．模擬会議システムを用いた3セットは使用する模擬会議システムの順番をグループ毎に変更した．被験者はFTFを含め合計4セット，12セッションの実験を行う．

3.3 アンケート

各セットが終了する度に被験者は表1にあるアンケートに回答する． 設問に対しリッカートスケールで回答する．

表1． アンケートの設問事項

設問1	数字のやりとりが自然にできた
設問2	無理なく発言することができた
設問3	タイミングの悪い発言が多かった
設問4	ぎこちないやりとりが多かった
設問5	不自然な沈黙が多かった
設問6	特定の一人に注目することができた
設問7	他の人が私に注意を払っているときがわかった
設問8	現在の数字についていくのが大変だった
設問9	不自然な動作が多かった

4. 分析

実験の様子は，全てビデオ録画した．録画データから発話のタイミングを4ミリ秒単位で書き起こした．全13グループのうち1グループは席の交換をしていないため、影響を考慮して分析対象から取り除き，残りの12グループを分析に利用した．

4.1 仮説1の検証

視線が伝達できないとき，意図しない相手が発言することを確認するため，矢印がある場合の発言ミスの回数を調べて，各実験設定に有意差があるか分析した．発言ミスした直前の数字に矢印がある場合を数え上げて，全発言ミスにおける矢印ありの比率を求めた．求めたデータに対して分散分析を行った．分散分析を行う際，比率データの正規化のため逆正弦変換を行った．

分析の結果，HydraとTVの設定で有意差を確認 した($F(1, 70) = 6.85$ $p<.01$)．同様にHydraとWebにおいても有意差があった ($F(1, 70) = 2.95$ $p<.05$)．また図4の矢印がある場合の発言ミスの比率の平均からHydraが他の設定に比べて発言ミスが少ないことが分かる．

4.2 仮説2の検証

仮説2を検証するため1回のタスクに掛けた時間(以降，タスク時間)を算出した．分析の中でノイズとなる3つの要素をそれぞれ補正した．以下，補正した内容について述べる．

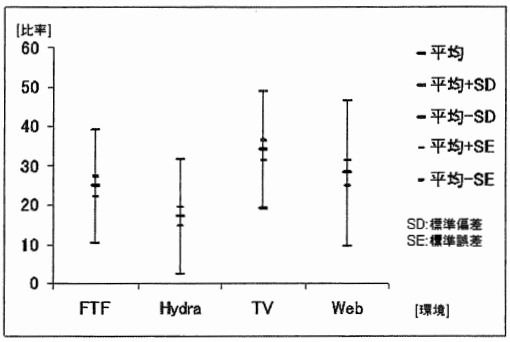

図4. 矢印がある場合の発言ミスの比率の平均

発言ミス時間の補正

グループによって発言ミスの数の偏りが大きく，それだけタスク時間に影響がある．従って，発言ミスした時間はタスク時間から引く形で除外した．数字を読み上げた時間から，次に間違った数字を読み上げた時間までが発言ミスした時間である．これにより，無駄な発言がなくなるため，タスク本来の掛かった時間が求まる．

難易度補正

タスクの数字と数字の差が大きいと，被験者は他の誰かがより小さな数字を持っている可能性を考えるため読み上げが慎重になる．本実験では，数字差が一定ではなかったために平均的に多く間違える問題があり，タスクの難易度が均等ではなかった．そのため，タスク毎に発言ミスの平均時間を算出し，その時間を難易度とした．この難易度に相当する時間をタスク時間から引くことでタスクの難易度を合わせた．

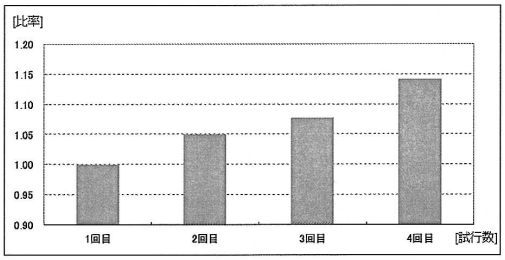

図5. 学習効果の影響

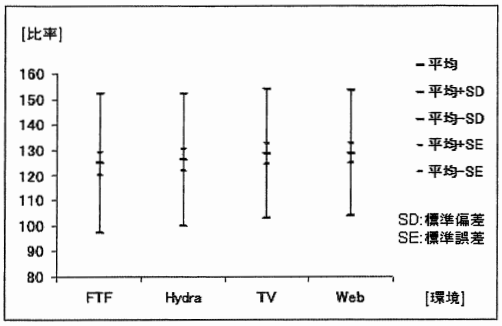

図 6. タスク時間の平均

学習効果の補正

補正による学習効果の排除を試みた. 図 5 は, 4つの設定で順番に行うとき, 2 から 4 回目のタスクの平均時間が 1 回目のタスクの平均時間に比べ, どの程度早くなったか, その増分を比率で表している. 図 5 から, 4 回目まで学習効果が落ちずに続いていることが分かる. この影響を取り除くため, 図 5 の増分の比率とタスク時間を掛けて補正した.

以上の補正を行なったタスク時間の平均値を図 6 に図示する. 図 6 から, 各設定において平均時間に差がなく, またバラツキも一様に大きい. この補正データに対して分散分析を行なった結果, 4設定全ての条件で有意差を得られなかった. 図 6 から, タスク時間に関しては FTF と Hydra は TV と Web に比べ 3, 4 秒早いが, 有意差を得られるほど早くないことを確認した.

4.3 仮説 3 の検証

TV と Web の設定において話者交替に差があるか仮説 1 で分析した内容に基づいて有意差を調べた. 4.1 節の仮説 1 の分析結果から, TV と Web に有意差を得られなかった($F(1, 70)=1.55$ $p>.05$). 従って, どちらの設定も話者交替において違いがなかったことが分かる.

4.4 アンケート分析

表 1 の各設問の得点に対して分散分析を行った. 表 2 にアンケートの平均得点を示す(表 2 カッコ内は標準偏差).

- 設問 1 は, Hydra と TV の間で有意差を確認した($F(1,188)=3.91$ $p<.01$). 被験者は TV よりも Hydra の設定においてスムーズにタスクを行えたと感じていることが分かった. また設問 2 も 5%有意水準で同様の傾向が見られた($F(1,188)=3.44$ $p<.05$).

表 2. アンケートの平均得点

	FTF	Hydra	TV	Web
1	4.12(1.6)	4.69(1.3)	3.83(1.4)	4.12(1.5)
2	4.77(1.5)	5.12(1.2)	4.37(1.4)	4.46(1.3)
3	3.87(1.6)	3.69(1.4)	4.18(1.2)	3.83(1.6)
4	4.33(1.6)	3.37(1.3)	4.35(1.4)	3.9(1.3)
5	4(1.8)	2.87(1.2)	3.4(1.6)	3.29(1.5)
6	3.1(1.8)	4.38(2)	3.15(1.7)	2.96(1.6)
7	4.71(1.6)	5.27(1.4)	3.92(1.7)	4.23(1.6)
8	3.17(1.7)	2.92(.1.7)	3.23(1.6)	3.1(1.6)
9	3.23(1.6)	2.79(1.4)	3.46(1.6)	3.04(1.5)

- 設問 4 は, Hydra と FTF および Hydra と TV の間で有意差を確認した(Hydra と FTF は $F(1,188)=4.34$ $p<.01$), Hydra と TV は($F(1,188)=5.03$ $p<.01$). これは FTF と TV にぎこちないやりとりが目立つことが分かった.
- 設問 5 は Hydra 以外の設定で不自然な沈黙が多いことが分かった (Hydra と FTF は($1,188)=6.70$ $p<.01$), Hydra と TV は($F(1,188)=4.84$ $p<.01$, Hydra と Web は($F(1,188)=7.59$ $p<.01$).
- 設問 6 は, Hydra の得点が有意に高く, 他の設定に比べ, 特定の人に注目できる(Hydra と FTF は ($1,188)=6.70$ $p<.01$), Hydra と TV は($F(1,188)=4.84$ $p<.01$, Hydra と Web は($F(1,188)=7.59$ $p<.01$)
- 設問 7 は, Hydra と TV および Hydra と Web で有意差を確認した. Hydra は他の設定に比べ, 視線の影響を感じやすいことが分かった. (Hydra と TV は($F(1,188)=9.10$ $p<.01$, Hydra と Web は ($F(1,188)=5.49$ $p<.01$)

5. 考察

Hydra と TV および Hydra と Web の間で, 発言ミス率に視線の有意な影響を確認できた. これにより, 仮説 1 が確かめられた. 一方, Hydra と同様に視線が伝達できる FTF に関しては, FTF と TV, FTF と Web の間で発言ミス率に対する視線の影響に有意な差を得られなかった. これは学習効果の排除に失敗したことが原因と考えられる. 本来は FTF 環境が最も視線を伝達しやすい環境である. しかしながら, 実験手順を説明するために, FTF での実験は常に最初に行っていたため, タスクの習熟が不十分であったと考えられる. 被験者がタスクに習熟するために, さらに練習セッションを設けるべきであったと考えられる.

タスクに要する時間はいずれの実験設定でも有意な差を確認できなかったため, 仮説 2 は確認できなかった. この原因として, タスクに使用した数字列

に難易度のバラツキがあった点が考えられる．直前に読み上げられた数字よりも一つだけ大きな数字を持っている場合には，確実に次は自分の番である．前の数字との差が大きくなればなるほど，自分の発言の番である確率は下がる．従って前の数字との差のバラツキがノイズとなったことが考えられる．この問題を解決するためには，２つの数字の差を一定の範囲内に制限し，セッション間の難易度を揃える工夫が必要であったと考える．

画質に関しては，Hydra と TV または Web の間では有意な差を確認できたが，TV と Web では有意な差がなかったことから，画質の違いで話者交替に有意な差がないことが確認できた．

アンケート分析では，被験者は，Hydra において，スムーズに話者交替できることが主観的評価で確認できたが，FTF では確認できなかった．ここでもやはり，タスクの習熟が不十分であった影響があらわれていると考えられる．

6. おわりに

本稿では，視線を伝達できるテレビ会議設定を用い，視線が話者交替にどのような影響を及ぼすか実験を行った．その結果，視線を正しく伝達できる場合，発言ミスの回数，つまり話者交替に失敗した回数が視線を正しく伝達できない場合に比べて有意に差があることが分かった．また視線の伝達の有無に関わらずタスクの時間に有意な差がないことを確認した．さらに画質の違いが話者交替に影響しないことが分かった．

今後は実験条件などを整え，さらに精度良く視線の効果をとらえられるよう工夫していく．

参考文献

[1] シード・プランニング，2008 年度版 テレビ会議/Web 会議の最新市場と HD 化動向，(2008/3).

[2] Argyle, M. and Cook, M., Gaze and mutual gaze, London: Cambridge University Press.

[3] Argyle, M. Bodily Communication, Routledge; 2 Edition, 1988.

[4] Kendon, A., Some functions of gaze-direction in social interaction. Acta Psychologica, 32, 1967, 1-25.

[5] Abigail J. Sellen, SPEECH PATERNS IN VIDEO-MEDIATED CONVERSATIONS, In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems(CHI '92), ACM Press, New York, 1992, 49-59.

[6] Vertegaal, R. and Van der Veer, G. and Vons, H., Effects of Gaze on Multiparty Mediated Communication, Proceedings of Graphics Interface 2000, 95—102, 2000.

[7] Vertegaal, R. and Slagter, R. and van der Veer, G. and Nijholt, A., Eye gaze patterns in conversations: there ismore to conversational agents than meets the eyes, Proceedings of the SIGCHI conference on Human factors in computing systems, 301-308, March 2001, Seattle, Washington, United States .

[8] Michael J. Taylor, Simon M. Rowe , Gaze Communication using Semantically Consistent Spaces, CHI 2000.

[9] David Nguyen, John Canny , MultiView: Spatially Faithful Group Video Conferencing, CHI 2005.

[10] Mukawa, N. and Oka, T. and Arai, K. and Yuasa, M. , What is Connected by Mutual Gaze? : User's Behavior in Video-mediated Communication , Conference on Human Factors in Computing Systems , 2005, 1677—1680.

[11] Ichikawa,, Yusuke and Okada,, Ken-ichi and Jeong,, Giseok and Tanaka,, Shunsuke and Matsushita,, Yutaka, MAJIC videoconferencing system: experiments, evaluation and improvement, ECSCW'95: Proceedings of the fourth conference on European Conference on Computer-Supported Cooperative Work , 279—292 , 1995.

[12] Kauff, P. and Schreer, O., An immersive 3D video-conferencing system using shared virtual team user environments, Proceedings of the 4th international conference on Collaborative virtual environments , 105—112, 2002.