

円滑な会話を実現する多人数ビデオチャット環境の構築

Eye Contactable Multi-party Video Conferencing System

佐藤 良 坂本孝丈 兵藤幸与 石川智之 市川 淳 後藤知彦 竹内勇剛
R. Sato, T. Sakamoto, S. Hyodo, T. Ishikawa, J. Ichikawa, T. Goto and Y. Takeuchi

静岡大学情報学部 認知的コミュニケーション研究室
cognitive, Faculty of Informatics, Shizuoka University
<http://cog.cs.inf.shizuoka.ac.jp>

1. はじめに

ビデオチャットは、直接会えない者同士が手軽に互いの顔・姿を映しながら語り合えるインターネットを利用したコミュニケーションアプリの1つである。近年は技術的進展に伴い1対1だけでなく、多人数間での会話環境も可能になってきている。ところが多人数会話（3人以上）になると、しばしば会話の円滑さが損なわれてしまう事態が生じる。

一般に3人以上での多人数会話を円滑に成り立たせるためには、次話者が誰になるかという選択が行なわれるのと合わせて、次話者として選択された者があるタイミングにおいて速やかに次話者としての順番の委譲を受けなくてはならない[1]. DuncanやKendonは、このような話者交替に伴う際に、視線や顔の向きが強く寄与している可能性を指摘しており、この考えは広く受け入れられている[2, 3]. このことから、円滑な多人数会話を実現するためには、会話参加者らの視線や顔の向きを参加者間で相互に参照し合える環境が要求されるはずである。

ところが一般的な多人数ビデオチャット環境では、会話参加者間で互いの姿を視覚的には参照し合えるのだが、話者が誰を次話者として選択して視線を向けているのか、あるいは誰が話者に対して次話者としての順番の委譲を要求しているのか、といった視線や顔の向きによる働き掛けを明確にするための配慮はなされていない。その結果、このようなビデオチャットを多人数で利用した場合、会話参加者間での発話開始時の重複が頻繁に生じるなど、話者交替が適切に行なわれず、会話の円滑さが妨げられるといった問題が顕在化している。

そこで本研究では、上述のような多人数ビデオチャット環境における視線による発話のアドレッシングや次話者のセレクト、次話者としての発話権の委譲要求が明示的に実行できないことによって妨げられる会話の円滑さを、新しいビデオチャット環境を提案することで改善することを目的とする。

具体的な取り組みとしては、会話参加者の視線動作を人物撮影用のカメラからの映像から解析し、会話参加者間で能動的なアイコンタクトを実現できる仕組みを従来の多人数ビデオチャット環境に実装する。このことによって、直接対面で行なわれる会話環境に近似したビデオチャット環境を実現することができ、複数の次発話者が発話冒頭に重複した発話を行なうなどの円滑な会話を妨げる原因をなくすことが期待される。さらに本研究で提案する多人数ビデオチャット環境は、特別な装置を必要とせずにご利用できたため、CMCにおける会話分析などのツールとして容易に利用することができる。

2. 自然な多人数ビデオチャット環境構築の試み

2.1 従来の多人数ビデオチャットの問題点

ビデオチャット(video conferencing)環境は、PCを端末として複数のユーザが互いの姿（主に顔部の映像）と音声相互に送受信するマルチモーダルなCMC(Computer Mediated Communication)である。かつては会話を目的としたCMCは文字によるIRC(Internet Relay Chat)が中心であり、後にIM(Instant Messenger)が普及し、その過程でファイルの送受信をする機能や、文字だけでなく音声による通話を実現する機能や、カメラで撮影した映像を音声と共に送受信する機能（→ビデオチャット機能）へと発展してきた。ビデオチャットは映像を取り扱うため通信路の帯域を多く消費し、端末での映像の描画処理が高負荷であることから1対1の会話環境に限られていたが、近年の技術発展によって多人数での会話環境も容易に実現できるようになってきた。

多人数ビデオチャット環境では、3人以上の会話参加者が互いの顔映像をそれぞれの端末の画面上に表示させるが、多くの場合では図1のように、会話参加者の顔映像をタイル状に並べる画面構成となっている。しかしこのような画面構成では、自分を中心とした1対1の会話環境が自分以外の会話



図1 多人数ビデオチャット環境の一例(1).

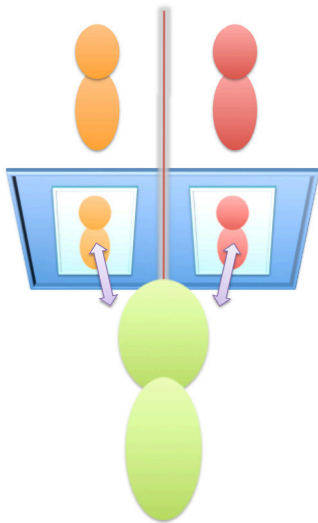


図2 1対1の会話環境が2つ同時に並行しているようなインタフェース.

参加者の人数分並列に提供されているような印象を与える(図2)。

一方、図3に示すような多人数ビデオチャット環境では、自分以外の会話参加者が円座になって互いの顔を参照しあっているようなインタフェースとなっている。このような映像の表示効果によって、会話参加者全員が1つの会話の場を共有しているような印象を各参加者に与えることができ、会話の活性化が期待される(図4)。

ビデオチャット環境は通常、PCに接続されたカメラで撮影された会話参加者の上半身映像(顔映像)が、他の参加者のディスプレイに表示される仕様になっている。カメラの多くはディスプレイ中央上部のフレーム内など、できる限り会話参加者がディスプレイを注視している際の顔を正面から撮影できる位置に設置される。そのため、各会話参加者のディスプレイ上には常に他の参加者が正面を向いた映像が表示されている状態で会話を行なう環境となっている。このような環境は、直接対面での会話場面では存在しえない。つまり一



図3 多人数ビデオチャット環境の一例(2).

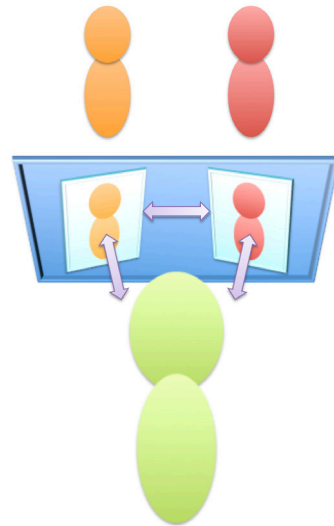


図4 3人が円座になっているようなインタフェース.

般的なビデオチャット環境では、会話参加者全員が互いに自分以外の参加者と正面を向き合っている状態となっているのである。

本稿冒頭でも述べたが、円滑な多人数会話を実現するためには、会話参加者らの視線や顔の向きを参加者間で相互に参照し合える環境が要求される。なぜならば、多人数での会話場面では、話者以外に複数の次話者候補者がいるため、次に挙げる項目が視線によって表現される可能性があるからである[4, 5]。

- (a) 話者が誰を次話者としてセレクトするのか。
- (b) 次話者としてアドレスされた者が次話者となることを受け入れるか。
- (c) 次話者としてセレクトされることを要求しているか。
- (d) 誰を次話者とすべきか。

しかし従来のビデオチャット環境では、話者やそれ以外の会話参加者の視線が、誰から誰へと向けられているかは適切に再現されない。したがって、上記のような視線によって表現される話者交替に関わるインタラクションは、従来のビデオチャット環境では成り立たない。その結果、(a)のような

場合には次話者を言語的に呼んだり、(c)のような場合には挙手をしたりすることで、視線以外のモダリティを用いた円滑な話者交替を試るほか有効な手段はない。

2.2 円滑な多人数ビデオチャットを実現するための先行研究

2.1節で述べたような視線による円滑な話者交替のためのインタラクションを、多人数で利用するビデオチャットシステムで実現するために、SellenらはHydra Unitと呼ぶ小型ディスプレイとカメラ、スピーカ、マイクを内蔵した装置を、自分以外の会話参加者の身体に見立ててテーブル上での多人数会話を試みた(図5)。この装置は、各参加者が個別に身体化されているため、ある装置に対する注視は、その装置に割り当てられている参加者への視線となり、直接対面した状況と近似した相互にアイコンタクトを可能にした会話環境を構築することができる[6]。

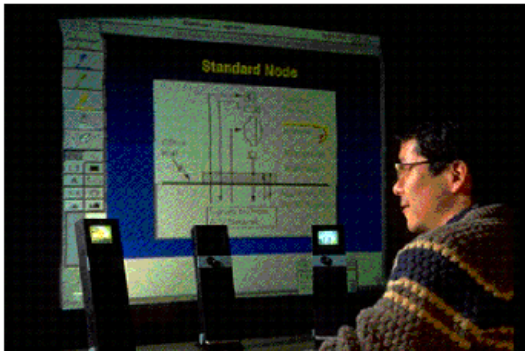


図5 Hydra Unitによる多人数会話環境。

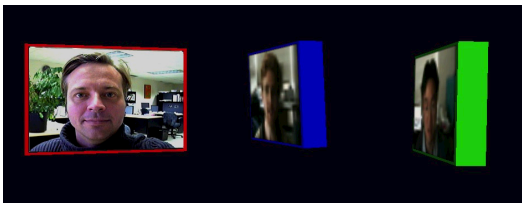


図6 Gaze-2による多人数会話環境。

またVertegaalらは、ハーフミラーを用いてハーフミラーの背後(会話参加者からは見えない)にカメラを複数台設置し、その各カメラとその会話参加者とを結ぶ線上のハーフミラー上に自分以外の会話参加者の映像を鏡面反射によって投影する装置によって、多人数でのビデオチャット環境におけるアイコンタクトの実現を試みた[7]。この装置はGaze-2と呼ばれ、各会話参加者は図6に示すように、自分が視線を向けている会話参加者の映像



図7 複合現実感技術を用いた多人数会話環境。

が正面を向くようにし、それ以外の会話参加者の映像は斜面像とすることで、誰が誰とアイコンタクトをしているかを明示的に表示させることができる[7]。

またこれに類似した多人数での共同作業環境として、OkadaらのMAJICなどがある[8]。さらに井上は、HMDによる複合現実感技術を用いることで複数の会話参加者を仮想的に1つのテーブルを囲んで会話しているかのような環境を、直接対面での多人数会話を行なっているかのような装置を通して、視線が会話の円滑な進行にどのような効果を与えているかを報告している(図7)[9, 10]。

2.3 提案する多人数ビデオチャット環境

2.2節で挙げたような、すでに開発された会話参加者間のアイコンタクトを可能とした多人数ビデオチャット環境は、SkypeやiChatAVのように一般に広く普及しているような市販されているUSBカメラをPCに接続し、プログラムを起動するだけの手軽さはない。

そこで筆者らは、より簡便に市販されている安価な機器と一般的なPCを使って、多人数の会話参加者が互いにアイコンタクトをすることができるビデオチャット環境の構築を試みる。

提案システムでは、図8に示すように、各会話参加者が仮想的に円形のテーブルを囲むような位置

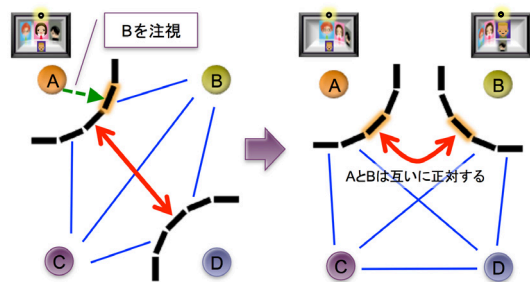


図8 視線の方向とアイコンタクト。

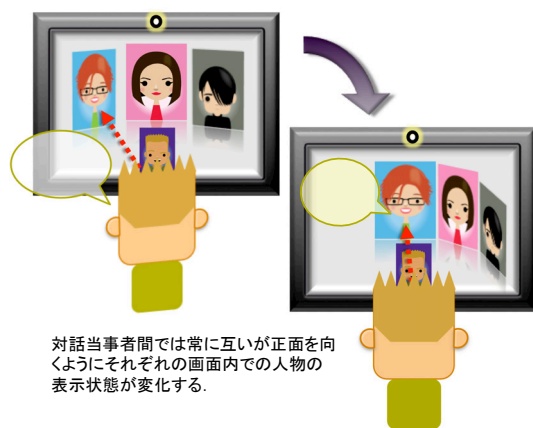


図9 会話参加者の映像の位置シフト。

関係に配置される。そして、図9のようにある会話参加者が視線を向けた人物の映像が正面に表示されるように、他の会話参加者の映像の位置が左右どちらかにシフトされる。また、他者から視線を向けられた会話参加者のディスプレイでは、自分を注視している者の映像がその位置で正対し、残りの会話参加者の映像が斜面像となることで、被注視状態として感じられるようになる(図10)。このとき、自分に正対している映像に視線を向けると、その映像がディスプレイ中央にシフトしてくる。これにより、2者間におけるアイコンタクトを能動的に行なうことができるようになる。ただし、会話参加者の正面に表示される映像は、あくまでもその者が視線を向けた他の会話参加者の映像となる。

会話参加者Pに視線を向けたSのディスプレイ

正面にPの映像がシフトしてくる。カメラと同一直線上で向き合う状態となる(図9)。

会話参加者Sに視線を向けられたPのディスプレイ

ディスプレイ上の位置はそのまま、Sの映像のみが正対し、残りの会話参加者の映像は斜面像となる(図10)。

本提案システムは先述した通り、特別な装置を利用することなく、一般的なユーザが通常利用しているようなコンフィグレーションで円滑な多人数ビデオチャット環境を構築するために、市販のUSBカメラ(Logicool製Qcam Orbit)から撮影した映像を、視線方向の検知にも利用する。そして各会話参加者が利用するPC(DELL製Studio XPS9000)上で、視線方向に応じた映像の表示位置を計算し、アイコンタクトをしている会話参加者同士が互いにディスプレイの正面にそれぞれの顔映像が表示されるようにソフトウェア的に操作している(図11)。

多人数ビデオチャットを実現する通信アプリケーションは、多人数会話(グループビデオ通話)に対

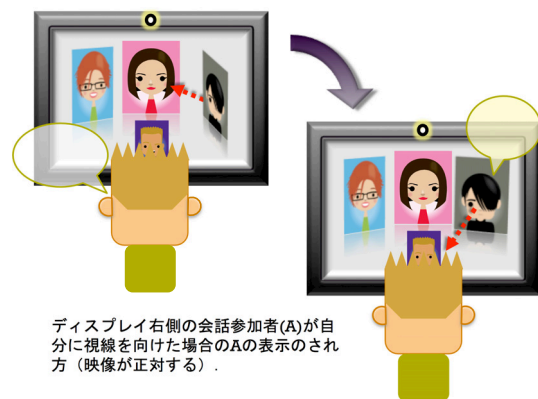


図10 自分を注視している会話参加者の表示のされ方。

応したSkype Premium 5.3 for Windowsを利用して、自分以外の会話参加者の映像がディスプレイ内で位置シフトする機構はOpenGLを用いて処理されており、ディスプレイに表示された映像のキャプチャし、Microsoft Visual C++でコーディングされた視線検知プログラムの出力に応じた位置に映像を再配置することで実現している。なお、視線検知にはOpenCVを用いており、左右への視線移動量を実時間で検出することができる。

3. 提案システムの評価実験

3.1 実験目的

本実験では、多人数ビデオチャット環境において、会話参加者の視線が円滑な会話の進行にどのような影響を与える要因となっているかを検証する。本実験においては、円滑な会話の進行とは次のような状態であると定義する。

- (A) 話者交替が行なわれる際の発話の重複が適切に回避されている。
- (B) 次話者が適切に順番の委譲を受けて発話を開始している。

このような円滑な会話が多人数間で成り立つためには、各会話参加者の視線の向きや動きが強く寄与していることが先行研究によってすでに実証されている。これらの知見は、直接対面している状況での多人数会話に基づいたものであり、ビデオチャット環境では、視線の向きや動きがどのように会話進行の円滑性に寄与するかは不明である。

本実験の仮説としては、ビデオチャット環境においても直接対面の状況と同様に、視線の向きや動きが多人数での円滑な会話を進行させていく上で必要な要因となると考える。したがって、本提案システムを利用した多人数ビデオチャット環境での会話は、会話参加者の視線の向きや動きを互いに参照可能であるため、直接対面での場合と同様

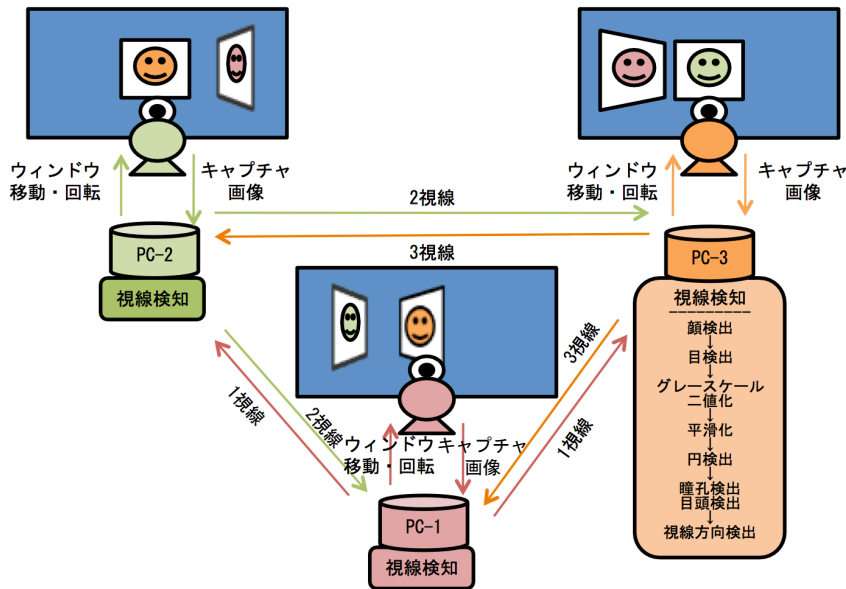


図 11 提案する多人数ビデオチャットシステムの概要.

な多人数会話の進行が観察されると予測する．同時に，本提案システムとは異なり，会話参加者間の視線の向きや動きを考慮していない一般に利用可能な多人数ビデオチャットシステムでは，本提案システムおよび直接対面の場合と比較して，円滑な会話の進行が妨げられると予測する（表1）．

表 1 会話環境と円滑な会話の進行に関する予測．

会話環境	視線の利用可能性	予測
直接対面	利用可能	円滑
提案システム	利用可能	円滑
従来システム	利用不可能	非円滑



図 12 直接対面条件での会話．

3.2 実験条件

本実験では，3.1節で比較したように，次の3つの実験条件のもとで次節で述べる観察指標に基づいた比較を行ない，本提案システムが円滑な多人数会話の進行に有効であることを実証する．

直接対面条件 図12のように各会話参加者が向かい合うように着座して会話をする．

提案システム条件 各会話参加者が，図13で示すような環境のもとでビデオチャットを行なう．

従来システム条件 各会話参加者が，図14で示すような環境のもとでビデオチャットを行なう．

本実験では，3人を1グループの単位として，5グループの実験協力者が会話参加者として各条件での多人数会話をそれぞれ1回ずつ体験する．グループ内の実験協力者の組合せは変わらない．

3.3 観察指標

本実験では，次の2つのデータを観察指標をしてデータの分析を行なう．

- (1) 発話開始時の重複（頻度）
- (2) 前発話と次発話との間（沈黙時間）

3.1節冒頭でも述べたように，本研究においては円滑な会話の進行は，話者交替が行なわれる際の発話の重複が適切に回避されていることと，次話者が適切に順番の委譲を受けて発話を開始していることであると定義している．

そのため，仮に発話の進行が円滑でなければ，発話開始時において複数の会話参加者による重複（同時に発話すること）が生じる頻度が高くなると予測できる．すなわち直接対面での会話の場合では視線の向きや動きによって話者交替が円滑に行なわれると仮定すれば，ビデオチャットの場合において，円滑な会話が行なわれているならば直接対面条件と同じ程度の頻度でしか発話開始時の



図13 提案システム条件での会話.



図14 従来システム条件での会話.

重複は発生しないと予測できる.

また次発話をすべき会話参加者が誰か不明な状態になり、次話者が決定せずに前発話と次発話との間に時間的な空白(沈黙時間)が生じてしまう可能性も予測できる. これに関しても、直接対面での会話の場合では視線の向きや動きによって話者交替が円滑に行なわれると仮定すれば、ビデオチャットの場合において、円滑な会話が行なわれているならば直接対面条件と同じ程度の間しか生じることはないと予測できる.

3.4 実験手続きと実験課題

本実験では多人数ビデオチャットにおいて円滑な話者交替が行なわれる際の視線の効果を検証するために、会話参加者の顔の向きや身振り・手振りなど視線以外の非言語的な表現を排除した形で実験を行なう. 実際には、図15のように、会話参加者としての実験協力者は、顔のサイズにくり抜かれたボードに顔をはめ込んで会話するため、視線の向きや動き以外に顔や体を左右に向けたリジェスチャーをしたりする非言語的な表現はできない環境になっている.

このような実験環境は、直接対面での会話においては違和感があるかもしれない. しかし、直接対面条件は他の2つの実験条件に対する統制条件になっているため、このような形にすることで、各条件間の結果の差異が視線の向きや動きを要因としていることが明確になる.

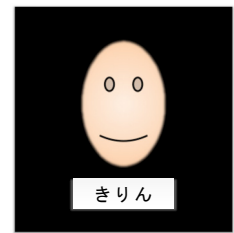


図15 会話参加者の正面像.

実験課題として、3名の会話参加者は次に述べる会話ゲームを行なう.

1. 3名の会話参加者は、それぞれ自分が顔をはめているボードの正面に書いてある動物や事物の名称を、他の参加者との会話によって誰よりも早く推理し、正解することを目指す(図15).
2. 会話参加者は、自分のボードに何と書いてあるかは知らない.
3. 会話参加者は、他の参加者を呼ぶ際にその者の名前を呼んではいけない. また、相手のボードに書いてある動物や事物の名称で呼んでもいけない.
4. 会話参加者は、積極的に他の参加者に自分のボードに書いてある名称に関するヒントを得るための会話を行なう. 会話参加者は基本的に虚偽の発言をしてはならず、ルールに抵触しないように可能な限り簡潔に答える.
5. 会話参加者は、互いに自分以外の参加者の推理を助けるための情報を言語的に提供する.

1ゲームは5分間とし、1つのグループは3つの実験条件のもとでの多人数会話を1回ずつ行なう.

3.5 実験結果

本実験は、3名1組のグループが3つの実験条件を1回ずつ行なうため、被験者内実験計画とみなすことができる. 実験では5組のグループが参加したが、実験時間の5分間でのグループ間での発話量はほぼ同程度であることが確認されている.

このことに基づいて、3つの条件間で観察指標である発話開始時の重複が発生した頻度を比較した(図16). データが5グループ分でしかないため統計的な信頼性が低いことは否めないが、1要因3水準での分散分析の結果、発話開始時の重複回数は会話環境の違いによって有意に差が生じることが示された($F_{(2,3)} = 9.71, p < .01$). LSD法による下位検定の結果は、従来システム条件が他の2つの実験条件よりも重複回数が多いことが明ら

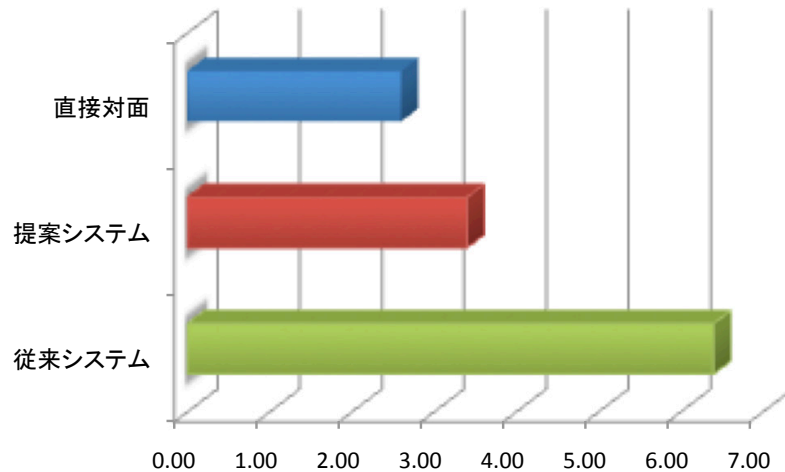


図 16 発話開始時の重複回数の条件間での比較 (回)。

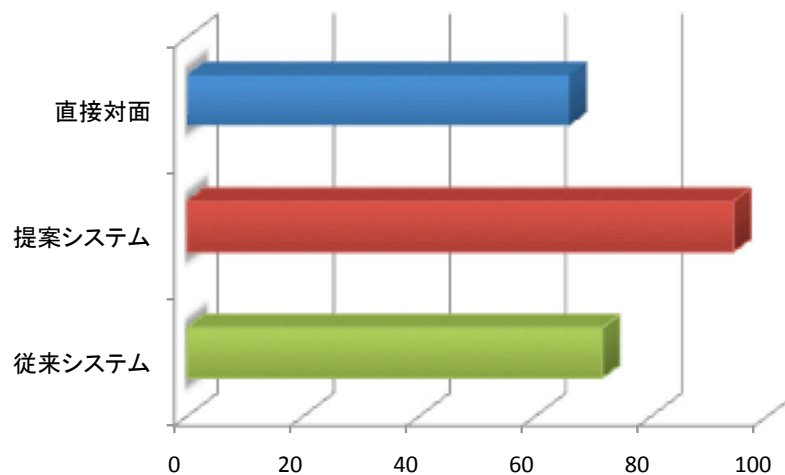


図 17 300秒のうちの沈黙時間の条件間の比較 (秒)。

かになった ($MSe = 2.07, p < .05$)。ただし、提案システム条件と従来システム条件の間での有意な差はなかった。

次に5分間 (300秒) の実験時間内に、3名の会話参加者のいずれもが発話していない状態の時間 (沈黙時間) の条件間での比較を行なった (図17)。その結果、発話開始時の重複回数の比較の場合と同様にサンプル数が不足しているため信頼性に乏しい統計値ではあるが、1要因3水準での分散分析の結果、沈黙時間は会話環境の違いによって有意に差はみられなかった ($F_{(2,3)} = 2.46, n.s.$)。

3.6 考察

本実験を通して、3.3節で示した次の2つの観察指標とそれに対する結果の予測の正否が明らかに

なった。

発話開始時の重複 (頻度) に関する予測

発話開始時の重複の頻度は、直接対面条件が最も少なく、それと同程度で提案システム条件が並び、従来システム条件はこの2つの条件と比較して高い値を示す。

前発話と次発話との間 (沈黙時間) に関する予測

沈黙時間は、直接対面条件が最も短く、それと同程度で提案システム条件が並び、従来システム条件はこの2つの条件と比較して長くなる。

まず、発話開始時の重複に関する予測は、図16が示すように、ほぼその正しさが検証されたといえる。一方、沈黙時間に関しては、図17が示すように、各条件間に差がなく、予測通りの結果にはならなかった。

このような結果となった原因として想定できることは次の通りである。

- (a) 提案システムを利用した場合には、会話参加者の視線の向きや動きによって自分の正面に表示される他の参加者の映像が切り替わるため、明示的な発話のアドレッシングや次話者のセレクトが行なわれ、発話開始時の重複が抑制された。
- (b) 提案システム条件では、会話参加者の視線の向きや動きに伴う映像の切り替えが発生するが、この際、会話参加者の視線の動きに適応できておらず、視線検知精度が一定でなかった。このため、会話参加者は映像の切り替えが完了するまで、前発話者にセレクトされた次発話者の発話を待ったことによって沈黙時間が増加してしまった。

上述の分析が示唆することは、提案システムが、本研究の目的である多人数ビデオチャット環境における視線による発話のアドレッシングや次話者のセレクト、次話者としての発話権の委譲要求が明示的に実行できないことによって妨げられる会話の円滑さを改善することに十分寄与できる能力を有しているということである。前発話と次発話との間の沈黙時間が、提案システム条件では直接対面条件と同程度にならなかったが、むしろこの結果は、提案システムが提供している会話参加者の視線の向きや動きに応じた他の参加者の映像の切り替えが有効に寄与したために生じたものであり、従来システム条件の場合の沈黙（誰が次話者か不明な状態で互いに発話を譲り合っている状態）とは異なる性質のものであると考えられる。

現時点では、データのサンプル数も不足しており、また分析および提案システムの視線検知の精度も十分なものではない。したがって本実験を通して、提案システムの有効性が実証されたという結論を導くことは尚早である。しかしながら、多人数ビデオチャット環境において、会話参加者の視線の向きや動きが会話の円滑な進行に大きく影響している可能性を示唆できた点で、本実験の成果はあったと考える。

4. まとめと今後の展望

本稿では次の点を中心に述べた。

1. 従来の多人数ビデオチャット環境では、会話参加者同士の視線の向きが誰から誰に向けられているか表現されず、次発話の開始時で発話が重複するなど円滑な会話の進行を妨げている可能性がある。
2. 会話参加者同士の視線の向きや動きを参加者間で互いに共有できる多人数ビデオチャット

環境を提案し、実験によってその有効性を検証した。

3. 多人数ビデオチャット環境では、会話参加者同士の視線の向きや動きが会話の円滑な進行に強い効果を与えることが示唆された。

3.6節で述べたように、本実験結果はまだ中途のものであり、今後より一層のデータの追加、および発話データ間の関係などの精度の高い分析と提案システムの改良が必要である。また、本システムを通して検知した視線の向きや動きのログデータをロボットの頭部の動きに反映させ、その際の会話音声と同期させることで、疑似的な直接対面での会話状況の再現を試みていきたい。

謝辞

本研究は、国立情報学研究所H21,22年度共同研究「多人数ビデオチャットにおける適切な発話アドレッシング手法の開発と円滑な対話環境の構築」（坊農真弓担当）による助成を受けている。

参考文献

- [1] Sacks, H., Schegloff, E. A., and Jefferson, G. (1974) "A simplest systematics for the organization of turn-taking for conversation," *Language*, Vol.50, No.4, pp.696-735.
- [2] Duncan, Jr., S. (1972) "Some signals and rules for taking speaking turns in conversation," *Journal of Personality and Social Psychology*, Vol.23, pp.283-292.
- [3] Kendon, A., (1967) "Some functions of gaze direction in social interaction," *Acta Psychologica*, Vol.26, pp.22-63.
- [4] 榎本美香, 伝康晴 (2003) "3人会話における参与役割の交替に関わる非言語行動の分析," *人工知能学会SIG-SLUD資料*, Vol.38, pp.25-30.
- [5] Lerner, G. H. (2003) "Selecting next speaker: The context-sensitive operation of a context-free organization," *Language in Society*, Vol.32, pp.177-201.
- [6] Sellen, A., Buxton, W., and Arnott, J. (1992) "Using spatial cues to improve video-conferencing," *Proceedings of CHI92*, pp.651-652.
- [7] Vertegaal, R., Weevers, I., and Sohn, C. (2002) "GAZE-2: an attentive video conferencing system," *Proceeding of CHI02*, pp.736-737.
- [8] Okada, K., Maeda, F., Ichikawa, Y., and Matsushita, Y. (1994) "Multiparty Videoconferencing in Virtual Social Distance: MAJIC Design," *Proceedings of ACM CSCW'94*, pp.385-393.
- [9] 野口康人, 井上智雄 (2007) "複合現実感を用いた分散会議における複数アバタの配置と表現," *情報処理学会論文誌*, Vol.48, No.1, pp.54-62.
- [10] 井上智雄 (2009) "実対人距離を調整可能な複合現実分散会議システム," *情報処理学会論文誌*, Vol.50, No.1, pp.246-253.