

近接チャットとHMD VRによるメタバースの相違点(ロング版)

長谷川 晶一 (Twitter: @hasevr) 東京工業大学 科学技術創成研究院

抄録- 本稿では、まず、オンラインのポスターセッション、懇親会、ミーティングを目的に開発した近接チャット (Proximity Chat = 平面マップを持ち音声の距離減衰機能を持つビデオチャット) である Binaural Meet の開発の背景、設計、機能を紹介する。次にその開発と利用を繰り返しつつ、メタバース (主に VR Chat) でのコミュニケーションと比べることで気づいたことを、会話分析、相互作用実践の研究を参照しつつ説明し、メタバースでのコミュニケーションの豊かさとその分析への期待を記す。

1. はじめに

2020年に始まったパンデミックは、我々の社会生活に大きな影響を与えた。筆者は2020年3月に開催された情報処理学会のシンポジウム「インタラクション2020」に携わっており[1]、会期15日前に急遽Zoomオンライン化を余儀なくされた。結果、口頭発表が問題なく進行する一方、デモ・ポスターセッションと懇親会は課題が残った。

ポスター・デモには1発表1Zoomを用意し、空いている部屋を可視化する意図で全Zoomの参加者をリアルタイムに一覧表示するWebページを用意した。結果としては、発表者だけの部屋へ入りにくいと考え参加者が多く、全く来訪者がいない発表が多数出てしまった。実世界のポスター発表では、発表者と直接会話する話し手や受け手にいきなりなるのではなく、最初は周りで話を聴き、理解を深め思いついたら質問をする話し手になるというプロセスがあるが、Zoom方式ではこれがうまく働かなかった。

懇親会は研究室ごとに行い、研究室間を1つのZoomで繋いだ。これにより他の研究室の雰囲気を知りながら研究室内で懇親を深めることは出来たが、研究室間の交流はほぼ教員の会話だけに限られていた。実世界の懇親会では、学生同士、教員同士の会話を局所的に行い、結果として多くの会話が並行して行われるが、Zoomでこれを再現することはできなかった。現在ではZoomもブレイクアウトルーム機能が充実したので、数人ずつランダムに割り振ったり、自由に移動可能にすることで、実世界の懇親会とは趣の違う懇親会を行うことはできるが、他の会話の様子がわからない点は変わっていない。

一方、バーチャル学会では2019年の第1回からポスター発表が現実のように機能している[2]。しかし、3Dのメタバース、特にHMD VR環境を2020年に学会や懇親会の参加希望者の全員に利用させることは難しいと考え、現在では近接チャットと呼ばれている2次元マップを持つビデオチャットをWebアプリとして実現すべくBinaural Meet[3]を開発した。

本発表では、近接チャットBinaural Meetを用いたポスター発表や懇親会、研究室ミーティングでの運用経験と、その後のVR Chat等のメタバースの利用経験を対比し、2D近接チャットとHMD VR環境のメタバースでのコミュニケーションの違いを考える。

2. 近接チャット

2.1 平面マップを持つ近接チャット

デスクトップPCやタブレット、スマートフォン等平面ディスプレイで使う場合、3次元の共有VR空間より、2次元の平面マップの方が表示と操作を単純にできる。そこで、3次元空間の代わりに2次元平面を共有することを考える。例えば、平面マップのロールプレイングゲームでも、キャラクターを平面マップ内で動かすことで様々な意図を表現できる。

ビデオチャットに共有平面をつけ、アバターを配置し、距離や位置関係等に応じて音量を減衰させると、複数の会話とそれを構成する参加者が存在でき、それらを可視化できるコミュニケーションツールとなる(図1)。さらに、Binaural

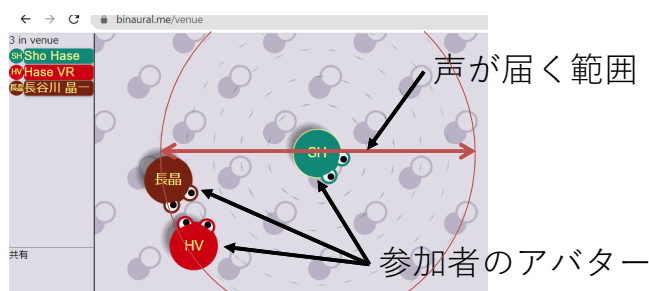


Fig. 1 近接チャット Binaural Meet のマップ、アバターと声の届く範囲の表現

Meet[3]やSkittisch[4]では、立体音響による音像定位機能を取り入れている。このような機能を持つWebアプリは2020年後半には多数開発・提供され、最近では近接チャット (Proximity Chat) と呼ばれている。

2.2 ブラウザによるビデオチャット=WebRTCの普及

IPでの音声伝送は1974年にまで遡る。77年には低遅延性のためにUDPができ、91年にパブリックドメインのVoIPソフトSpeak Freelyがリリース、96年のRTPのRFCと続く。これをWebブラウザに載せ、ルータの内側のブラウザ同士をできるだけ効率の良い通信経路を探してつなぐ仕組みを加えたのがWebRTCであり、2011年から標準化が始まり、音声映像の圧縮方式と合わせて発展し、主要なブラウザで互換性が取れたのが2019年と、丁度パンデミック前に普及が進んだ。

2.3 近接チャットの提供状況

近接チャットのサーベイ論文を見つけることはできなかったが、パンデミック以降に発表された近接チャットの紹介し

ている修士論文 [5] や記事 [6] がある。McCulloch の記事 [6] は、2020 年 11 月時点として 42 の URL が並ぶ Simpson と Zuegel のリスト [7] を紹介しているが、2022 年 10 月現在 Web ページがなくなっている URL も多数ある。

近接チャットの多くは、オンラインの懇親会やオフィスを意図しているが、パンデミック以前の 2010 年から、複数の部屋を持つ平面マップを共有するバーチャルオフィスを意図した Web 会議 Sococo [8] が提供されている。他にもパンデミック前からの取り組み [9, 10] がある。上記のリストには掲載されていないが、現在では、日本の企業もツールを提供しており [11, 12]、国内で普及しつつあるように見える。

3. Binaural Meet とその設計

筆者は、インタラクション 2020 での経験から近接チャットの必要性を感じ Binaural Meet[3] の開発をはじめた。2020 年 7 月頃から研究室でのオンライン懇親会に使いはじめ、2020 年 12 月には研究室でのミーティングもこの上でやっている。近接チャット機能に加え、共有平面内に作業に必要な機能をもたせている。ポスターセッションや懇親会での利用も意図しているが、研究室のミーティングでは毎日のように利用している (図 2)。また、大学院の講義でのディスカッション (図 3) やポスター発表 (図 4) でも利用した。

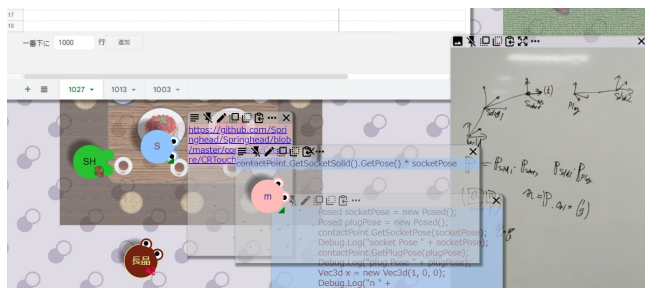


Fig. 2 Binaural Meet を用いた研究室ミーティング

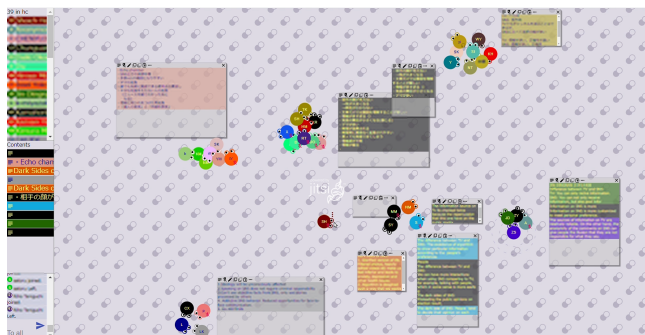


Fig. 3 Binaural Meet を用いたグループディスカッション

3.1 発話者の表現

多人数会話であり、平面上にアバターが存在するため、誰がどこで発話しているかが分かる必要がある。アバターの発声音量をアバターを囲む点線の円で表示する (図 1) とともに、Web Audio API を使った音像定位により発話者の位置を表現する。音像は、自己のアバターからの相対方向で定位させる方式と、画面上を前方、下を手前、左右を左右に対応さ

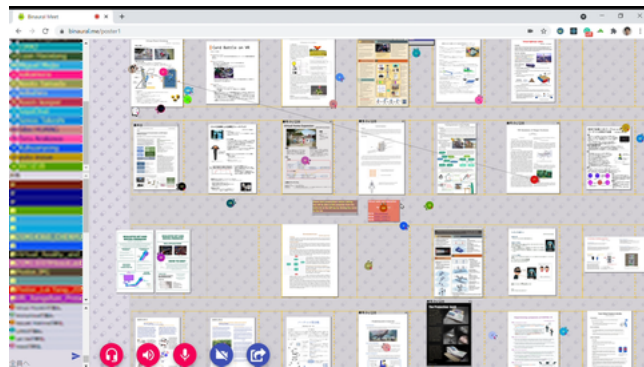


Fig. 4 Binaural Meet を用いたポスター発表

せることで利用者視点で定位させる方式を選択できるようにしたが、何方を選んでもあまり判らないという利用者も居た。定位を活用する利用者はアバター相対を好む者が多い。また、音像定位を利用すると音質が良くなるという印象を持つ利用者が多かった。ただし、Chromium では Web Audio API を使うとエコーキャンセリングが機能しないため、Chrome と Edge では、ヘッドフォンの利用が必須となる。

3.2 研究室ミーティングのための共有機能

画面や画像、動画だけでなく、付箋的なテキストや共有ホワイトボード WBO[13] を共有できる。また HTML の iframe 要素を利用して Google Docs を共有し、共有平面内での共同作業を実現している。共有コンテンツは誰でもアレンジ (移動と拡大縮小) できるので、協力して並べ替えたり、複数の参加者の画面共有を並べて議論したりと行った使い方ができる。

3.3 イベントの性格とアクセス権

学会では、自分のポスターを見張っていなくてもポスターが剥がされる心配はない。オンラインでも同等と考え、誰でも画像や PDF 等をアレンジ・削除できるようにした。運営者と参加者の立場が近く、全員で作るイベントでの協同作業を意図してこのような設計とした。これまでの経験では、まれに操作ミスによる削除等の事故があるがすぐ修復できている。誰でも掲示やアレンジができることから、運営側の手間を少なくできていると思う。

3.4 会話参加者の変化を促す仕組み

研究室のミーティングでは教員と学生だけでなく、学生同士の小声でのローカルな会話により知識が補完、共有される。また、視線や体の向きなどで、誰がどの会話に興味を持っているのかを判断しながらコミュニケーションが進められる。これらは距離減衰だけでは再現が難しいため、次のような機能を付加した。

■糸電話機能 会話の輪の参加者のうち特定の人だけに話しかける耳打ちは距離減衰だけでは実現しにくい。実世界ほどのダイナミックレンジや音質がなく、平面のアバターでは耳に口を近づけることが難しいため、耳打ちを再現し難いためである。そこで、特定の人だけに声を届ける糸電話機能を用意した。

■**会議室機能** 距離減衰が常に働くと、多くの参加者が一つの会話に入る場合に、スペース不足からアバター同士が重なってしまい参加者が把握できなくなる。これを解消するために、会議室＝距離減衰が働かない矩形のスペースを用意した。

3.5 アバターの向きと視線

利用者がアバターを操作しなくても、利用者の興味を他の利用者に伝えることができるようにするため、マウスカーソルと画面中央をアバターの視線と身体の向きで表現した。共有コンテンツを見る場合、画面のスクロールと拡大縮小によりコンテンツをウィンドウの中央に大きく写すユーザーが多い。そこで、画面操作時にアバターの身体をウィンドウ中央向きになるように回転させ、興味の対象を表現するようにした。また、マウスカーソルはより瞬間的な興味を反映するので、アバターに目を持たせ、視線をマウスカーソルに向けるようにした。

3.6 3D アバター

Kalidokit[14] と three-vmr[15] を用いることで、Web カメラの映像から顔や全身のランドマークを MediaPipe[16] を用いて検出し、頭部や全身の姿勢を求め、VRM アバターに適用して Web ブラウザ上に表示することができる。そこで、Binaural Meet 標準のアバターの上に VRM アバターを載せるられるようにした (図 5) ブラウザ上 three.js での描画のためポリゴン数が多くなると描画が遅くなる。また MediaPipe による全身の認識は計算負荷が高いため、顔のみ検出しポリゴン数の少ないアバターを利用している。VRM アバターの場合でも体の向きは前述のように興味の対象を向くようにし、頭部や四肢のみを利用者の動きに連動させた。



Fig. 5 Binaural Meet のアバター。左から名前のイニシャル、ビデオ、顔特徴点、顔トラッキングによる VRM、全身トラッキングによる VRM。顔特徴点では、顔の向きをアバターの向きに傾きを目と鼻の位置にカメラからのユーザの距離をアバターの前後移動に対応させた。VRM では顔の位置と向きまたは全身の姿勢にアバターが連動する。

4. VR Chat と Binaural Meet の比較

本節では、非言語コミュニケーションの研究を参照しながら、VR Chat では見られるが Binaural Meet では見られないコミュニケーション方法を主に指摘する。

4.1 非言語コミュニケーション

Binaural Meet のアバターはマウスとキーボードで操作でき、位置向き 3 自由度と注視点 2 自由度の計 5 自由度を持つ。しかし、注視以外はキーボードを用いるかマウスでアバター

をドラッグする必要があるためか相槌等のジェスチャには使われていない。一方、VR Chat では頭と手の 3 点 × 6 自由度やフルトラッキングのアバターを用いる利用も多い。このため、非言語コミュニケーションに大きな差が出るが予想される。

非言語コミュニケーションの研究は、Darwin の人と動物の感情表現の研究 [17] の後、Efron[18, 19]、Birdwhistell[20, 21] による記述・分類・文化間の比較がなされ、Ekman[22] は Efron の分類を継承しつつジェスチャを emblem, illustrator, affect display, regulator, adaptor に分類している。VR Chat でもバイバイと手を振るジェスチャ (emblem) や位置関係や大きさを表すジェスチャ (illustrator) が使われる。感情表現の affect displays は顔表情については少し特殊で、顔表情の計測装置があまり普及しておらず、手指のジェスチャと顔表情を連動させ手指で表情を操作している利用が多い。傾き等会話の調整をする regulator、状況や感情に合わせて無意識に起こる adaptor も VR Chat でも手や頭部の動きについては見られる。注視の計測もあまり普及していないため、眼球運動は再現されず、頭部動作だけが注視の手がかりとなる場合が多い。Kendon[23] は Efron[19] Ekman[22] を含むこれまでの分類による理論がそれぞれ役立つが完全ではなく様々な視点からの研究が必要だと指摘している。また、発話を助けるジェスチャーと発話と独立なジェスチャーは分けられるものではなく、連続だと指摘している。そして、録画から書き起こした発話とジェスチャの時系列の記録から、発話とジェスチャの関係や様々なジェスチャを分析している。これらがメタバースでどのように生じているか (生じていないか) を知るためには、メタバースでのコミュニケーションの際に行われる発話とそれ以外の行為の分析が必要となるだろう。このような会話や会話が始まる前からのコミュニケーションで人々が行う発話を含む行為は、相互行為実践 (Interactional Practices) と呼ばれ、次節の話者交代や陣形を含め、研究されている。小室と船越 [24] はロボットとの会話を分析し、特有の会話構造があること、ロボットに対する相互行為実践が人に対するものと異なることを指摘した上で、人らしさの評価につながる可能性を示している。この手法は、メタバースのアバターと実世界の人がどれほど近く、また違うのかを知るためにも利用できるかもしれない。

Binaural Meet の利用では、話者から質問があった場合に Yes と答えるための傾きとしてマウスを動かすことでアバターの眼球を動かす者が多いが、アバターを動かしたり、画面をドラッグすることでアバターの向きを変える者は見かけない。また、話者の語りに傾くためにマウスを動かすこともあまりない。一方で Binaural Meet でも 3D アバター機能を用いると、傾きなどの頭部運動がそのままアバターに反映される。研究室ミーティングで筆者が低ポリゴンの 3D アバター [25] を使うと、研究進捗報告の際に話が理解されたかどうか分かり話しやすいという意見を、複数の学生から得てい

る。しかし、多人数が3D アバターを使うと利用環境によっては処理落ちや切断等が生じるため、多数の利用者が3D アバターを用いた場合の行動は観察できてない。顔特徴点のアバターでも頭部運動に応じてアバターが動くが、動きが小さく分かりにくいいためか、あまり利用されていない。以上は筆者の印象であり、より具体的に生じている現象を知るためには、具体的な会話での行為の時系列を記録し分析することが必用だと思う。

4.2 話者交代

多人数会話では、次に誰が発話するかを決めることと、いつ話者が交代するかが決まる必要がある。Sacks ら [26] は複数の聞き手から次の話者を指定するには、現在の話者は予め呼びかけや視線を向けてある聞き手に伝えておく必要があると指摘している。また、Kendon[27] は、発話の終わりで現話者が次話者となる聞き手に視線を向け、聞き手が話者に視線を向けることが話者交代の合図になるとしている。現時点のメタバースでは視線は再現されず顔の向きだけが再現されることが多いが、顔の向きが話者交代の合図として使われているかどうかを筆者は判断できておらず、記録と分析が必用だと思う。

Binaural Meet では、3D アバターを用いた場合でも第三者視点になることから、視線や顔の向きによりある聞き手を示すことは難しい。

4.3 F 陣形

Kendon [28, 29] は、制約のない空間で多人数が会話するときに人々が陣形を形成すること発見し、F 陣形 (F-formation、F は Face の意と本人が答えたらしい [30]) と名付けた。F 陣形は会話する人々の体の位置と向きにより作られる。生き物の活動には活動のための空間が必要であり、例えば TV を見る場合は TV までの錐台の空間が必要となる。この空間は操作領域 (transactional segment) と呼ばれ、会話の場合には参加者がこれを共有するために F 陣形ができると考えられる (図 6)。Kendon は会話が行われる中心のスペースを O 空間 (o-space)、その外側の参加者の立ち位置を P 空間、その外側を R 空間と呼び、O 空間が保たれることや、新たな参加者は P 空間に近づくが R 空間に留まり、会話の参加者が気づき促すまで待つこと等を指摘している。VR Chat でも多くの雑談で F 陣形が作られる。F 陣形は円形とは限らず、例えば野球観戦しながらの横並びの例も示されている [29]。

一方、Binaural Meet では、例えば図 3 のように O 空間が作られない。これは共有平面内のアバターにはそもそも操作領域が存在しないためかもしれない。常用している筆者や研究室メンバーも、他者のアバターが自己のアバターの目の前を横切ってもなにか用かな？と思う程度でそれほど邪魔には感じない。一方で画面共有などマップ上の作業対象の上にアバターがあると邪魔に感じる。これは、操作領域が画面平面内ではなく、画面平面内の作業対象 (画面共有) と利用者の間の実空間に存在するためだと考えられる (図 7)。

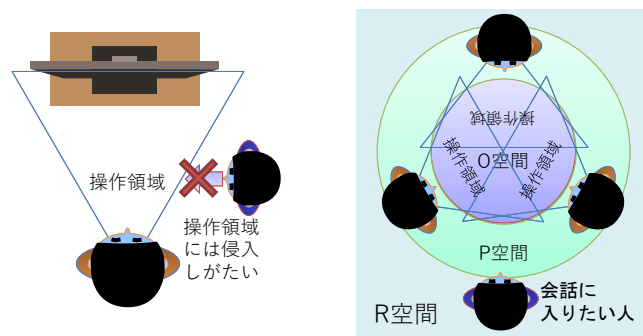


Fig. 6 操作領域と F 陣形

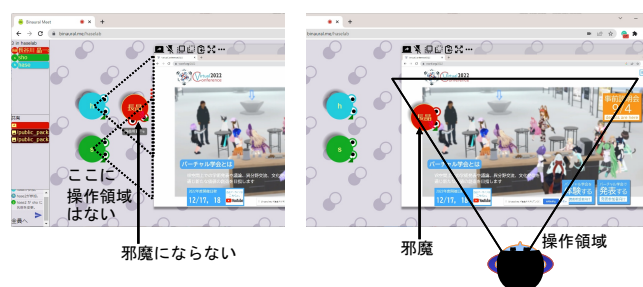


Fig. 7 平面近接チャットの場合の操作領域

また、研究室メンバーからの聞き取りで、Binaural Meet のアバターの軽薄さを示す次のような意見が得られた。

- VR Chat だと会話に入らないでその場に居ることが辛い。3つの会話の輪の中央に居ながらどれにも入ることができず、また聞きとることも難しかったので、退出してしまった。Binaural Meet の場合は、ミュートで居続けられる。Zoom のカメラ Off に近い。
- VR Chat も透明なアバターがあればいられるかもしれない。HMD をかぶっていて他の事ができないこともあるかもしれない。

4.4 Web ブラウザでは標準だがメタバースではできていないこと

Web ブラウザでは iframe を用いることで他のサービスを埋め込むことができる。Binaural Meet でもホワイトボード [13]、YouTube、Google Docs 等を埋め込んでいる。一方、現在のメタバースでは、コンテンツはサービス提供者のサーバだけから提供されており、例えば VRoid Hub のアバターギャラリーを VR Chat の World に埋め込むことはできない。

もともと Web は双方向リンクを捨て単方向リンクにすることで、情報提供を行う Web サーバとその管理を情報提供者に任せ、編集権限を持たない Web サーバにリンクできるようにしたことで発展した。メタバースにもこの仕組みを踏襲する日がきて、様々なサービスを埋め込めるようになることを期待したい。

5. 終わりに

本稿では、近接チャット Binaural Meet 開発の背景とその設計と機能を紹介した後、会話分析、相互行為実践の研究を参照しつつ、VR Chat では見られるが、Binaural Meet では

見られないコミュニケーションの行為を、筆者の気づいた範囲で紹介した。また、Web では簡単な他のサービスの埋め込みがメタバースではできていないことを指摘した。

期待と実態の区別がつきにくいメタバースだが、メタバースの大きな価値の一つは人に会えること、より端的には会って雑談ができることだと思う。筆者は2次元の近接チャットを作り運用することで、逆にメタバースでのコミュニケーションの豊かさに気付かされた。会話分析、相互行為実践の研究は、このメタバースの価値をより詳細に示すことや、より簡易ながらメタバースと同等の効果をもつバーチャルメタバースの設計のヒントになると期待している。

謝辞

筆者は、相互行為実践という言葉で2022年春に初めて知り、メタバースでのコミュニケーションの面白さにより自覚的になることができた。末筆ながら、相互行為実践の研究を紹介いただき、学内研究プロジェクトにお誘いくださった東京工業大学の船越孝太郎氏、会話分析、相互行為実践についてお話しした千葉大学の小室允人氏、早稲田大学の牧野遼作氏に深く感謝する。小室氏にはVR ChatでF陣形が作られることをご指摘頂いた。

また、メタバースでのコミュニケーションが近接チャットと大きく違うことを指摘し、一度メタバースを見てくることを強く勧めてくださった、明治大学の三武裕玄先生に深く感謝する。最後に、近接チャットBinaural Meetの開発に協力くださったGuo Zile氏、Wang Jingge氏、Li Haoyan氏を始め、日常的に利用しフィードバックくださる研究室メンバーに深く感謝する。

参考文献

- [1] 長谷川晶一, 志築文太郎, and 小坂崇之. インタラクシオン2020オンライン開催の舞台裏とそこから見えたこと. *情報処理学会論文誌*, 62, 2021.
- [2] Live#218【#バーチャル学会2019】ポスターセッション:vr空間からvr世界へ / 公式配信. YouTube <https://www.youtube.com/watch?v=bQdvQs713nY&t=5355s>, 2022年10月アクセス.
- [3] Guo Zile, Haoyan Li, Jingge Wang, and Shoichi Hasegawa. Binaural Meet. <https://binaural.me/>, 2022年10月アクセス.
- [4] Skittish. <https://skittish.com/>, 2022年10月アクセス.
- [5] Benjamin Maus. Designing video call spaces with and for adults with learning disabilities: A remote participatory design approach. Master's thesis, Malmö University, 2021.
- [6] Gretchen McCulloch. A mission to make virtual parties actually fun. <https://www.wired.com/story/zoom-parties-proximity-chat/>, 2022年10月アクセス.
- [7] Star Simpson and Devon Zuegel. list of proximity chat platforms as of mid-nov 2020. <https://docs.google.com/spreadsheets/d/1FgynDnnrrl1ZZUa42XcKSfD68G5VAMZsLkARXHtgHOM/edit?usp=sharing>, 2022年10月アクセス.
- [8] Sococo LLC. Sococo - smart spaces to communicate naturally. <https://web.archive.org/web/20101229144334/https://www.sococo.com/home.php>, 2022年10月アクセス, 12 2010.
- [9] John Palmer. Spatial interfaces. <https://darkblueheaven.com/spatialinterfaces/>, 2022年10月アクセス.
- [10] FunCorp Lab. Spatial Chat. <https://spatial.chat/>, 2020年9月アクセス.
- [11] ovice. <https://ovice.in/ja/>, 2022年10月アクセス.
- [12] NTT Communications. NeWork. <https://newwork.app/>, 2020年9月アクセス.
- [13] Wbo—共同作業できるオンラインホワイトボード. <https://wbo.ophir.dev/>, 2022年10月アクセス.
- [14] Rich (yeemachine) 他. Kalidokit. <https://github.com/yeemachine/kalidokit>, 2022年10月アクセス.
- [15] 0b5vr 他. @pixiv/three-VRM. <https://github.com/pixiv/three-VRM>, 2022年10月アクセス.
- [16] Kinar R (kinaryml) 他. MediaPipe. <https://github.io/mediapipe/>, 2022年11月アクセス.
- [17] Charles Robert Darwin. *The expression of the emotions in man and animals*. John Murray, London.
- [18] David Efron. *Gesture and Environment*. King's Crown Press, New York.
- [19] David Efron. *Gesture, Race and Culture (Reissue of Efron 1941)*. Mouton and Co., The Hague.
- [20] Ray L Birdwhistell. *Introduction to kinesics: (An annotation system for analysis of body motion and gesture)*. Department of State, Foreign Service Institute, 1952.
- [21] Ray L Birdwhistell. *Kinesics and context: Essays on body motion communication*. 1970.
- [22] Paul Ekman and Wallace V Friesen. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *semiotica*, 1(1):49–98, 1969.
- [23] Adam Kendon. *Gesture: Visible action as utterance*. Cambridge University Press, 2004.
- [24] 小室 允人 and 船越 孝太郎. 相互行為実践としての対話ロボットに対する態度. *人工知能学会論文誌*, 37(1):A–L61.1–15, 2022.
- [25] メシエナンバー. 256chinchilla/brown. <https://hub.vroid.com/>, 2022年10月アクセス.
- [26] Harvey Sacks, Emanuel A Schegloff, and Gail Jeffer-

- son. A simplest systematics for the organization of turn taking for conversation.
- [27] Adam Kendon. Some functions of gaze-direction in social interaction. *Acta psychologica*, 26:22–63, 1967.
 - [28] Adam. Kendon. *Conducting interaction : Patterns of behavior in Focused Encounters*, volume 7. Cambridge University Press, 1990.
 - [29] Adam Kendon. *Spacing and Orientation in Co-present Interaction*, pages 1–15. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.
 - [30] 坊農 真弓 and 高梨 克也. **多人数インタラクションの分析手法 (知の科学)**. オーム社, 2009.