



対面 - 遠隔ハイブリッド型会議において COVID-19 感染対策と円滑な音声対話を両立させる試行実験

An Experimental Study of Designing a System to Achieve Both COVID-19 Infection Control and Smooth Voice Dialogue for in-Person and Remote Hybrid Meetings

瓜生大輔¹⁾, 澤田怜旺²⁾, 柏野善大¹⁾, 稲見昌彦¹⁾

Daisuke URIU, Reo SAWADA, Zendai KASHINO and Masahiko INAMI

1) 東京大学 先端科学技術研究センター (〒 153-8904 東京都目黒区駒場 4-6-1, uriu@star.rcast.u-tokyo.ac.jp)

2) 早稲田大学 先進理工学部応用物理学科

概要: 同室内に複数人が集い場とその他の遠隔参加者を交えたいいわゆる「対面 - 遠隔ハイブリッド型会議」では、物理的距離の確保や飛沫拡散防止対策を行いながら全参加者の発話をクリアに集音・配信するための工夫が求められる。本研究では、このような環境下における円滑な音声対話を実現するための音響システムを東京大学内の研究室に設置し、連続的な試用と改善を行ったプロセスを報告する。

キーワード: テレプレゼンス, 音声コミュニケーション, 遠隔会議, COVID-19, 感染症対策

1. 序論

COVID-19 感染拡大によるリモートワークの一般化は、個人々が遠隔会議用機器を準備する契機となった。会議の出席者全員がそれぞれ別の場所から接続する場合は、各人が PC や適切なスマートフォン用のマイクやヘッドセットなどを準備することで円滑な会議進行が可能である。しかし、一部の出席者たちがオフィスなどに集い、遠隔 - 対面のハイブリッド型会議となると状況が一変する。もともと遠隔会議専用の設備が導入されていないオフィスの場合、とりわけ音声を明瞭に集音・配信するための工夫が必要となる [1]。対面での飛沫感染が懸念される中、十分な距離を確保しながらすべての発言を明瞭に集音することの両立には困難が伴う。1つの会議用マイク・スピーカーの周りに複数人が密に集まることは望ましくない。複数人が同一空間内で複数のマイクを起動すると機器ごとの通信遅延差により音声重畳し聞き取りづらくなる他、エコーキャンセリング機能によりまったく会話が成立しない場合もある。

本研究では、このような状況への解決策の一つとして、通常ライブハウスなどに導入されている PA (Public Address) システム¹ をミーティングスペースに導入し、遠隔会議用途への応用を試みる。数十人、数百人が一同に会す会議場の場合、大音量で出力するための PA システムが用いられるが、今回の試みはあくまで 4~5 人が対面で集い、その他多数の人が遠隔参加するような状況を想定する。さらに、発

話者同士の距離があり、あるいは飛沫感染を防ぐための仕切りがあり、発話が聞き取りづらい状態でもある。これらを考慮しつつ、それぞれの対面会議参加者に適切な音場空間を提供するために本研究を着想した。

本研究は、飛沫感染予防という課題から着想されたものではあるが、テレプレゼンス研究における音場空間設計手法例を、そしてコロナ禍以後も継続すると考えられるハイブリッド型会議を円滑にかつ高品質に遂行するための知見を提供するものである。今日のテレプレゼンス研究は、レイグジスタンス研究に代表される視覚・聴覚以外の感覚伝送に関するもの [2] や、全方位カメラなどを用いた視覚経験の拡張 [3] などに注目が集まっている。一方で、音声を題材とした研究については立体音響技術などを用いた没入感や臨場感の向上などに主眼を置く研究が見られる [4] が、あくまで脇役的な位置づけに過ぎない。本研究成果は、他システムとの比較や詳しい性能評価などを含むものではなく、他所での再現性を保証するものではない。しかし、「多拠点接続型遠隔会議における音声制御」という、今後も避けては通れない課題を提起するとともに、一定の知見を提供するものである。

2. 研究手法

2.1 音響システムの基本構成

本研究の舞台は東京大学先端科学技術研究センター内の Living Lab Komaba に設置された直径 2,400mm の掘りごたつ型の円形テーブルを中心とする縦 3m、横 5m ほどのミーティングスペースである (図 1)。約 8 名ほどが同時に座ることができるが、COVID-19 感染対策として、現在は定員を 4 名に限定している。アクリル板で仕切られた区画および 1.5m ほど離れたソファをそれぞれ区画①~⑤と表記

¹PA 用のミキサーは、複数のマイクや楽器からの音声入力を適切な音量に調整し、ライブハウス会場内などに出力する機材である。これに加え、演奏者ごとに個別にミックス音声を作り出力する AUX 出力機能がある。これは、ステージモニターなどに使用される出力で、(観客に向けた)メインミックスと独立して(各演奏者の要求に応じて)個別のミックス音声を出力できる。

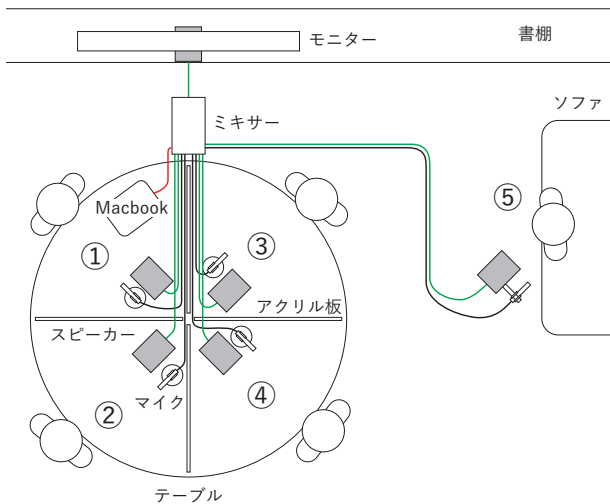


図 1: 音響システムの基本構成

する。各区画には集音マイク (SAMSON C02) とアクティブ・モニタースピーカー (YAMAHA MSP3) を配置し、ミキサー (Soundcraft Ui24R) に接続している。SAMSON C02 はステレオ集音用のペア・コンデンサーマイクであるが、高い指向性と音質を兼ね備えたコストパフォーマンスの良い機材であるため採用した。Soundcraft Ui24R はすべての操作を WiFi 経由で行える、オーディオインターフェース機能を兼ね備えたデジタル・ミキサーである。20 本のマイク入力に加え、8 本の AUX 出力を搭載する。①に配置された遠隔会議用マシン Apple Macbook は、ミキサーからの音声を遠隔会議システム zoom に USB 接続で送信する。zoom から出力される (遠隔参加者からの) 音声は、ステレオ・フォンケーブル経由でミキサーに出力される。遠隔会議時には Macbook から HDMI 出力や Apple TV などを通してテーブル前方の書棚に置かれたモニターに zoom の画面が表示される。①～⑤の基本構成に加え、モニターの下にはもう 1 台スピーカーが配置され、ミキサーと接続されている。こちらには zoom から出力される (遠隔参加者からの) 音声のみが出力される。図 1 に描かれた黒色のケーブルはマイク入力 (+ 48V ファンタム電源) 緑色のケーブルは AUX 出力を表す。

2.2 研究プロセスの概要

本研究は、上記ハードウェア設定を維持した状態で、2021 年 5 月から 7 月にかけて実際にハイブリッド型会議を行いながら音声の入出力に関する諸設定の改善を繰り返すことにより遂行した。具体的には、Soundcraft Ui24R のデジタルミキサー上で、各マイクからの入力音量を操作する GAIN、①～⑤それぞれのスピーカーへの AUX 出力音量の調整、特定の音声周波数の音量を調整するイコライザー (以下 EQ)、設定した閾値より小さい入力音をカットする GATE、大きな入力音を抑制し出力音量差を低減するコンプレッサー (以下 COMP) を調節することにより、音響の改善を図った。

約 2ヶ月の研究期間、ミーティングテーブルでは平均週 2



(a) ①への AUX 出力 GAIN 設定



(b) ①からの入力への EQ 設定



(c) ①からの入力への COMP 設定

図 2: AUX 出力 GAIN, EQ, COMP の初期設定

～3 回、1～2 時間程度の会議が開催された。改善プロセスは、著者らが各会議に同席した際に主観的に把握した問題点と、テーブル上に配置した質問紙、Google フォーム、研究室内の Slack チャンネルなどに記入された使用者からのフィードバックを反映する形で遂行した。また可能な場合は zoom の録画機能を用いて、適宜、会議終了後にも音声を分析した。

3. システムの改善プロセス

3.1 初期設定

研究開始当初、すべてのマイクの GAIN を約 10dB とし、各スピーカーへの出力音量は AUX 出力設定で調整した。①～④の各 AUX 出力は、該当区画に配置されたマイクからの入力音<以外>を一律約 5dB 出力する設定とした (図 2a)。他区画から距離がある⑤のみ約 6dB とした (図 7a)。これは、他の発話者からの発声音のみ一律に増幅することを意図した設定である。EQ は会話を強調するために 2～5kHz 前後の周波数を強め、こもった音として聞こえる 500Hz 周辺を抑制し、可聴域以外の高低音をカットした (図 2b)。ま



図 3: ①の GATE 設定

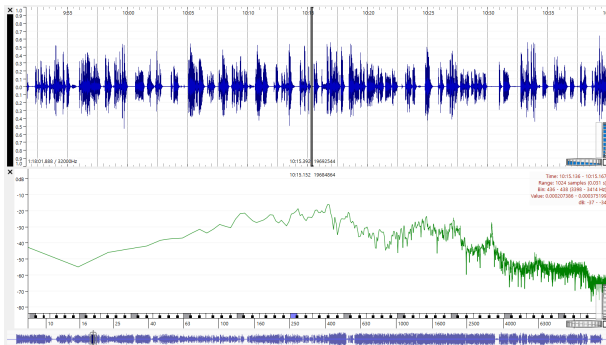


図 4: ハウリング時の周波数スペクトル

た、マイクから集音された音声の出力音量差を抑えるために図 2c の通り COMP を設定した。GATE は使用せず、閾値 0dB とした。以後、これらの設定を変更・改善したプロセスについて述べる。

3.2 改善プロセス 1

初期設定時、時折、高周波の残響音をともなうハウリングが発生した。この問題に対処するため、改善プロセス 1 では、GATE 及び EQ の調整を行った。GATE の設定では、各区画において発話がなされていない場合に環境音が集音されないように閾値を設定した。研究室という空間の特性上、空間内には様々な音源が存在する。空調機器をはじめ、近くにあるシンクの水道、電子レンジなどによる騒音を測定した²。しかし、GATE の閾値を上げすぎると、騒音・環境音のみならず発話音もカットされ、会議参加者に不快感を与える可能性がある。以上を考慮し試行錯誤した結果、GATE の閾値を騒音の最大値に対して 15dB 程度抑える設定とした(図 3)。これに加え、ハウリングを引き起こす周波数を特定し、EQ の調整を試みた。ハウリングが確認された会議の録画音声の波形分析を行った結果、3kHz 付近の周波数で耳につく残響音が確認された(図 4)。これをふまえ、初期設定時より 3kHz 付近の周波数を抑制した(図 5)。

3.3 改善プロセス 2

本システム構築の動機のひとつに、アクリル板により遮られた状態でありながら、距離が離れた区画同士の発話を聞き取りやすくすることがある。このためには可能な限り出

²無指向性マイク (MXL V67N) をテーブル中央高さ 1 m ほどの位置に配置して、集音・測定・分析した。



図 5: ①の EQ 設定

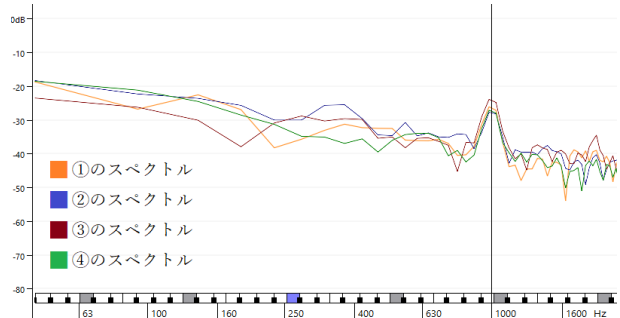


図 6: ②から 1000Hz の音声を出力した時の各区画から測定した周波数スペクトル

力の GAIN を上げる必要があるが、その分ハウリングが発生しやすくなるトレードオフがある。改善プロセス 2 では、とりわけ声が届きにくい向かい合う区画(①と④、②と③)同士の出力 GAIN を上げられるか検討した。まず、向かい合う区画同士のみ、初期設定から 1dB GAIN を上げ 6.0dB とした。しかしながら、会議の最中にハウリングが発生した。会議後に 5.5dB に設定した状態でハウリングが生じないか、音量が適切であるかを確認するために 100Hz、250Hz、500Hz、750Hz、1000Hz、2000Hz、3000Hz の音を③、④から発生させ、周波数スペクトルを観察した。この結果、ハウリングは生じず、各区画から聞こえる音量差も小さいことがわかった。入力音量と周波数スペクトルを同時に観察したところ、ハウリング時に起きる「特定の周波数帯のみが徐々に上昇する」現象は観測されなかった。一方、音量を各区画から計測するとすべて約 10dB 内に収まっていることが確認された(例えば、図 6 は、③から 1000Hz の音を出したときの各区画で計測したスペクトルを表す)³。以後、実際の会議中に問題点が発生しないか検証を続けた。

3.4 改善プロセス 3

他区画(①~④)と離れた⑤には、独自の設定を行った。とりわけ①、②、④との距離があり、発話が聞き取りづらいため調整を施した。具体的には、AUX 出力 GAIN を①からは 7.9dB、③からは 5.9dB に設定した。これは筆者の 1 人が実際に⑤から会議に参加した際、①、③からの参加者による発話を聴きながら主観的に判断し設定した。変更

³試験音には各周波数のサイン波を用い、各区画の話者着席位置から無指向性マイク (MXL V67N) で計測した。



(a) 初期設定



(b) 改良後設定

図 7: ⑤への AUX 出力 GAIN 設定

により、比較的距離がある①からの発話も聞き取りやすくなった。一方、初期設定では遠隔参加者からの発話はモニター下のスピーカーからのみ聞こえるようになっていたが、⑤からは聞き取りづらかった。zoom 経由の遠隔参加者からの発話音声 -8.6dB の GAIN 設定で出力することによりこの問題を改善した。

4. 課題と今後の展望

改善プロセス 3 までを終え、頻繁に発生していたハウリングの頻度を低減させ、また出力 GAIN を上げることで初期設定時よりも会話を聞き取りやすい音場を構築することに成功した。しかしながら、複数人が同時に生活する研究室の特性上、軽微な変化がシステムのパフォーマンスに影響する。例えば、各区画に置かれたマイクやスピーカーは完全には固定されていないため、機器の特性を理解している使用者が、適切な向きに動かしたり、口元に近づける、といった行動が見られた。一方で、そもそもシステムが稼働していることに全く気づかない使用者も存在した。無意識にシステムの恩恵を受けられる場合にはむしろ理想的ともいえるが、前使用者がマイクやスピーカーの位置を極端に移動させていた場合には問題となる。今後、本システムを改善する場合にも、使用者によって音響機器使用リテラシーに差があることを考慮にいれなければならないだろう。

使用者から得られたフィードバックの中に、「⑤から初めて発話があった際、どこから聞こえたのかわからず戸惑った」という意見があった。円卓を囲んでいる者同士であれば視覚的に誰が発話しているかを認識でき、生音も聞こえ

るため定位感もともなう。しかし、比較的距離のある⑤からの発話は、遠隔参加者の音声なのか、同一空間内からの発声なのか区別できなかったと考えられる。現在、各区画はモノラルで集音・発音しているが、ステレオ化や立体音響の導入など [4] で改善が見込まれる。しかし、音源が増えれば増えるほど、その扱いは複雑化する。現場の目的に見合った改善案の検討が必要であろう。

システム使用時には、zoom を経由した遠隔参加者からのフィードバックも適宜得られた。幸い、ほとんどの会議において音声は明瞭に聞き取れたが、発話の頭尾が切れるあるいは、唐突に切れた感じがする (GATE の効き過ぎ)、一時的にノイズが目立ったといった指摘があった。前者については GATE の微調整によりほぼ改善するが、その一方で環境音に起因するノイズについては除去に限界がある。オフィスや研究室といった日常的な環境におけるハイブリッド会議は今後も一定の頻度で重用される [1] ことを展望すると、音響工学の専門家でなくとも容易に使いこなせる環境音ノイズ対策手法開発は一つの課題といえるだろう。

本稿では、物理的な距離を取りながら複数人が対面する一方で多数の遠隔参加者とともに「ハイブリッド型会議」のための音声 PA システムの構築と改善プロセスを報告した。本研究成果は一研究室室内での限られた試みであり、普遍的な理論を提起するものではないが、日常生活空間におけるハイブリッド会議のための音場空間構築という課題への一対処法として、今後の研究に応用・反映したい。

謝辞 本研究は JST ERATO (JPMJER1701) からの助成を受けたものである。

参考文献

- [1] Barco Co., Ltd: “What are hybrid meetings? Why should you prepare your offices for them?”, in <https://www.barco.com/en/news/2020-11-03-what-are-hybrid-meetings> (2020)
- [2] S. Tachi, Y. Inoue, F. Kato: “Telesar VI: Telexistence surrogate anthropomorphic robot VI”, *International Journal of Humanoid Robotics*, Vol.17, No.05, p.2050019 (2020)
- [3] M. Manabe, D. Uriu, T. Funatsu, A. Izumihara, T. Yazaki, I. Chen, Y. Liao, K. Liu, J. Ko, Z. Kashino, A. Hiyama, M. Inami: “Exploring in the city with your personal guide: Design and user study of T-Leap, a telepresence system”, in 19th International Conference on Mobile and Ubiquitous Multimedia, pp.96–106 (2020)
- [4] I. Salselas, R. Penha, G. Bernardes: “Sound design inducing attention in the context of audiovisual immersive environments”, *Personal and Ubiquitous Computing* (2020)