複数 Web 会議に同時参加するための映像交互倍速再生・音声字幕化システム

山本貴文[1,a)]　土田修平[1,b)]　寺田 努[1,c)]　塚本昌彦[1,d)]

概要：Web 会議は場所による制約を受けずにリアルタイムでのコミュニケーションが可能なため，二つの会議に同時に参加できる．しかし，複数の Web 会議に同時に参加する場合，複数の会議内容をリアルタイムに理解できないため，質問や意見などの発言が困難となる．そこで本研究では，複数の Web 会議に同時に参加している状況において，参加しているすべての Web 会議の内容を理解し，発言などのインタラクション支援するシステムの構築をする．提案システムは録画された二つの会議映像の再生速度を上げることにより再生時間の短縮を行う．それらの映像を短い間隔で交互に視聴することで複数の Web 会議への同時参加を試みる．また，会議音声の字幕化，タイムシフト再生を用いて会議内容の理解を支援する．本稿では，提案システムを実装し，同時刻に行われる二つの Web 会議への同時参加か可能となるか調査した．結果，ユーザが複数の Web 会議内容の理解することを支援できたが，一方，リアルタイムの会議への発言や応答に 20 秒程度の遅れが生じ，ユーザ以外の会議参加者に違和感を与える結果になった．

1. はじめに

働き方改革や新型コロナウィルスの感染拡大防止などにより，Web 会議システムは急速に普及している．従来の対面による会議と比べ，人と人同士の接触を避けることにより，コロナウィルスの主な感染要因とされている飛沫感染や接触感染 [1] を防ぐことができる．また，会議場所への移動にかかる時間と交通費を削減でき，資料や会議室の確保といった事前準備にかかる手間とコストを抑えることができる．よって，業務効率化とウィルスの感染予防の観点において，Web 会議の需要と重要性は高まっている．また，ネット環境と PC といった最低限の環境を整えることで，場所による制約を受けないという特徴をもつ．そのため，Web 会議システムを使用すれば二つの会議に同時に参加することが可能である．

しかし，複数の Web 会議に同時に参加する場合，会議内容に対する質問や意見などの発言が困難となる．その理由として，主に二つの課題がある．一つ目は，会議内容をリアルタイムに適切に理解することが難しい点である．複数の会議映像を同時に視聴して，議題の異なる会議内容を理解することは困難である．高田ら [2] は，2.2 倍速の二つの遠隔会議映像を一部重複しながら 44 秒間隔で交互に視聴することで同時参加を支援したが，リアルタイムの会議から数十秒遅れるため円滑なインタラクションは実現できていない．二つ目は，ツールに関する課題である．複数の Web 会議に同時に参加する機能は既存の Web 会議ツールに無く，複数ツールの操作を会議の進行と同時に行うことは困難である．これら問題の解決に向けて，同時に参加しているすべての Web 会議への発言や返答を行うために，会議の内容理解を支援する手法を検討する必要があるが，筆者らの知る限りこれまで調査されてこなかった．

そこで本研究では，複数の Web 会議に同時に参加している状況において，参加しているすべての Web 会議の内容を理解し，発言などのインタラクションを支援することを目的とする．本稿では，先行研究 [3] で提案した二つの Web 会議の内容を理解を支援するシステムをもとに，二つの Web 会議へのインタラクションを支援するシステムを実装し，有効性を調査した．

2. 関連研究

2.1 音声・映像を高速で視聴する研究

音声や映像メディアの内容を短い時間で理解するための研究は数多く行われている．長濱ら [4] は，1 倍速，1.5 倍速，2 倍速の再生速度の異なる映像コンテンツの学習効果を調査した．理解度テストの分析結果から，2 倍速までの再生速度の違いは学習効果に影響を与えないことが示唆さ

[1] 神戸大学大学院工学研究科
Grad. School of Engineering, Kobe University
[a)] takafumi-yamamoto@stu.kobe-u.ac.jp
[b)] t.sway.tmpp@gmail.com
[c)] tsutomu@eedept.kobe-u.ac.jp
[d)] tuka@kobe-u.ac.jp

れた．しかし，2 倍速に対するアンケート調査では高速再生による疲労感が指摘されており，認知負荷が高まる可能性がある．栗原 [5] は，映画などの字幕付き動画の鑑賞方法として，主映像と言語情報を独立して制御する変則再生システム CinemaGazer を提案した．また，音声からの情報理解を諦めることで鑑賞時間を平均 85.5%削減できることが示された．また粥川ら [6] は，ウェアラブルカメラで撮影された一人称視点の映像において，重要な部分を強調しつつ，高速で閲覧するためのインターフェース DO-Scanning を提案している．DO-Scanning は，画像認識を利用して映像中の物体を分類し，強調箇所の候補とする．ユーザは各候補の中から物体の重要度を設定することにより，重要な物体が映ったシーンは元の速さで，他のシーンは高速で再生することで，映像全体を高速で閲覧する．Lee ら [7] は，人物に着目して一人称視点の映像を自動要約するシステムを提案した．このシステムでは，映像に映る人物や，一人称視点の映像を用いて，重要なシーンを検出し，冗長なシーンをカットした．別のアプローチで視聴時間を削減する研究として，中野ら [8] は，インターネット上の動画共有サイトに投稿されている動画の複数同時視聴を支援する Web アプリケーションを提案した．このシステムでは，ユーザが Web サイトから視聴したい動画を選択すると，動画プレーヤーがアプリケーション上に配置される．この操作を繰り返すことで，1 画面上に複数の動画プレーヤーが配置され，同時視聴ができる．

このように音声や映像を高速で視聴するための研究は数多く行われているが，これらは情報の高速受容を目的としており，リアルタイム性は考慮されていない．本研究では，リアルタイム性を考慮した，会議映像の内容理解を支援するシステムを検討する．

2.2 会議の要約に関する研究

会議内容を要約し，会議参加者への適切な情報提示を検討する研究は多く行われている．Hsueh ら [9] は，目標を達成するために，複数の手段や代替案から最適なものを選ぶ意思決定に着目した自動要約ブラウザを開発した．これは，会議の結論とその過程に関する情報をモデルにより推定する．従来の要約よりも，議論での意思決定に関連する記録を効率的に見つけることを可能にした．また会議の途中参加を想定した要約システムとして，水田ら [10] は，議論内容の摘要を用いて，途中参加の支援を行うテキストベースの電子会議システムを開発した．このシステムでは，会議全体の流れをまとめたものと，議題内容ごとにまとめたものを提供することで，途中参加者が参加以前の議論情報を取得することを支援している．このシステムを利用した結果，会議の全発言ログの利用頻度が低下したことから，ダイジェストが議論内容の把握を代替したことを示唆している．他にも Tuker ら [11] は，録音された会議音声の重要な部分のみを抽出・再生することで，会議への途中参加を支援するシステム catchup を開発した．これは TF-IDF を用いて会議音声中に出現する単語の重要度をスコア化することで，抽出する部分を決定している．Shi ら [12] も会議音声データを用いて，過去に参加した会議の要約を図を用いてビジュアル化することで，テキスト表示よりも直感的な要約を提供する Meeting Vis を開発した．Meeting Vis では，タイムライン形式で，話し合われたトピック，頻出したキーワード，発言が活発であった人を表示することによって，会議の詳細情報の把握を支援する．James ら [13] は，議題，発言者の役割といった事前情報と音声認識を用いて，仮想会議の要約を行う V-ROOM を提案し，要約の品質を向上させた．

しかし，これらの要約は自動で行われるため，ユーザにとって重要である議論を見逃してしまう可能性がある．本研究では，音声認識技術を用いた会議音声の字幕化とユーザが任意の会議映像を見返すことで，会議内容の理解を支援する．

2.3 遠隔会議の同時参加を検討する研究

遠隔会議の同時参加を検討する研究は多数行われている．安西ら [14] は，重複して流れる二つの音声に対する内容理解ついて調査した．この研究では，二つの音声をそれぞれ交互に聞く場合と同時に聞く場合で，内容理解度に大きな差はないと報告している．しかし，同時に聞く場合では，正確に内容を理解できているか確信を持てない被験者がいた．また高田ら [2], [15] は，二つの遠隔会議映像の短縮再生と映像切り替えを用いてリアルタイムに近い時間で会議内容を理解し，そのうち一つの遠隔会議への参加を支援している．会議 A と会議 B の二つの会議映像を蓄積し，それらの映像を 2.2 倍速再生することで再生時間の短縮を行う．また，44 秒間隔で映像をスイッチングすることで，リアルタイムに近い時間で二つの映像を交互に視聴する．しかし，ユーザが早送りの会議映像で見た場面はリアルタイムでは数十秒前の場面であり，そこから会議への発言を行っても，会議の進行に対してスムーズな参加とは言えない．

これらの研究では二つの遠隔会議の内容理解を検討しているが，双方のインタラクションについては検討されていない．本研究では，同時に参加しているすべての Web 会議へのインタラクションを検討する．

3. システム設計

本研究の想定環境を図 1 に示す．参加者はそれぞれ異なるメンバーで構成されており，トピックも異なる，二つの Web 会議に同時に参加する状況を想定している．このような状況において，参加している Web 会議内のユーザの発言や返答を通常の会議と同様に行うことができれば，複数

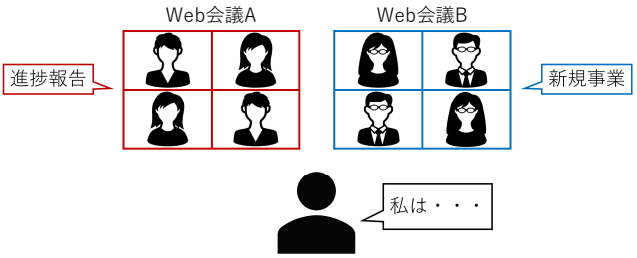

図 1 想定環境

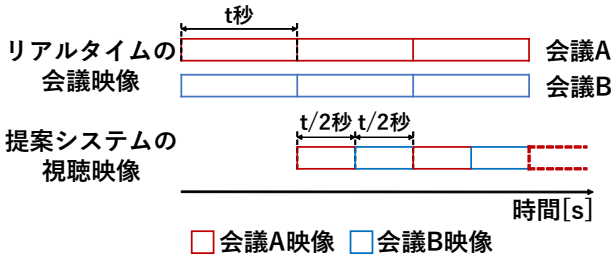

図 2 視聴イメージ

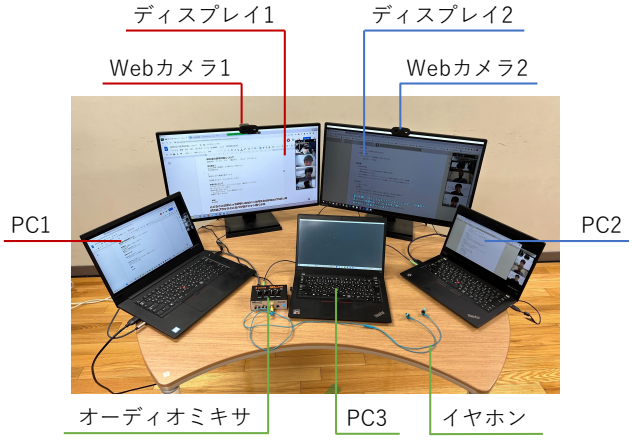

図 3 デバイス構成

の Web 会議に同時に参加できたと言える．本章では，以上を達成するシステムの構築に向けて，満たすべきシステム要件と必要な機能について述べる．

3.1 システム要件

複数の会議映像を同時に視聴する場合，互いに音声が打ち消し合うため，トピックの異なる会議内容を理解することは困難である．同時に視聴をしない場合，通常速度で再生を行うとリアルタイムとの遅れが生じる．また，Zoom [16] や Microsoft Teams [17] のような Web 会議ツールを用いる場合，複数のツールを同時に操作する必要がある．会議内容を聞き取りながら，マイクミュートの切り替えなどの操作を行う必要がある．以上より，以下二つの項目をシステム要件とした．リアルタイムから遅れずに会議内容を理解できる．発言を行う操作が簡易である．

3.2 機能設計

前節のシステム要件を踏まえて，問題点を解決するシステムの機能について述べる．まず，二つの会議映像をリアルタイムに取り込み，映像の再生速度を上げ，それら二つの会議映像を短い間隔で交互に視聴する．通常速度で映像を視聴する場合，二つの映像を視聴する時間が必要となるため，視聴時間に応じてリアルタイムからの遅れが大きくなる．そこで，会議映像の 2.0 倍速で視聴することで，映像視聴に必要な時間を短くする．視聴イメージを図 2 に示す．先行研究 [2] では，2.2 倍速の二つの遠隔会議映像を一部重複しながら 44 秒間隔で交互に再生することで同時参加を支援していたが，本研究では映像が切り替わる間隔をさらに短くし，リアルタイムとの遅れをより短くする．通常，再生速度を上げると音程も高くなるため，内容理解が困難となる可能性がある．ピッチを変えず，音声の再生速度を上げることは内容理解にとって有効であるため [18]，Phase Vocoder と呼ばれる周波数領域で音程や時間といった情報を制御するアルゴリズム [19] を用いて，音程を変えずに再生速度を変える．また，音声認識技術を用いて，会議音声データからリアルタイムで文字起こしを行い，会議の字幕を作成する．会議の発言をすべて文字起こしすることで，視覚情報による内容理解の補助を行う．さらに，視聴映像の表示に加え，マイクのオン・オフといった Web 会議ツールの操作をキーボードで行うことで，発言を行う操作を容易にする．

4. 実装

実装したシステムのデバイス構成を図 3 に示す．システムは PC を 3 台使用し，Web 会議音声・映像を処理・再生する PC1，PC2 と，PC1 と PC2 を通信・操作する PC3 から構成される．1 台の PC で使用できるシステムが望ましいが，実装簡略化のため PC を 3 台使用した．PC1 と PC2 で処理された映像をそれぞれ映すため，ディスプレイを 2 台使用する．また，音声の出力制御はオーディオミキサを使用し，一つのイヤホンで会議音声を聞く．2 つの Web 会議にユーザを映すためのカメラ，マイクとして Web カメラを 2 台使用する．PC は Lenovo 社の ThinkPad を 3 台，ディスプレイは I・O DATA 社の 27 型ディスプレイを 2 台，オーディオミキサは Donner 社の 4 入力 2 出力オーディオミキサ，Web カメラは Logicool 社の C505e HD BUSINESS WEBCAM を使用する．

4.1 システム構成

システム構成を図 4 に示す．PC1 とディスプレイ 1，Web カメラ 1，オーディオミキサはそれぞれ接続されており，PC2 も対応するデバイスと同様に接続されている．PC1，PC2 と PC3 は OSC（Open Sound Control）と呼ばれるデバイス同士でデータの送受信を行う通信プロトコル

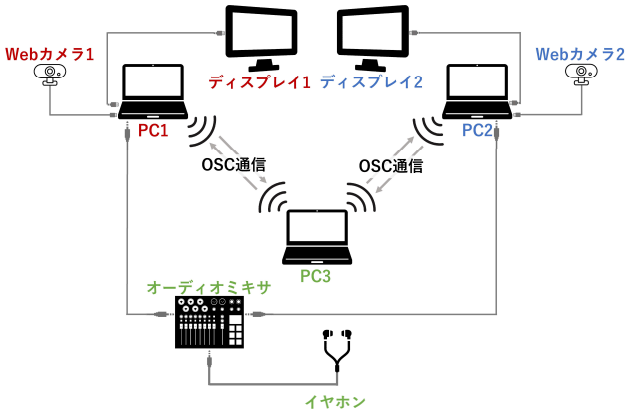

図 4　システム構成

図 5　PC アプリケーション構成

を用いて接続する．また，PC1 と PC2 で再生される音声はオーディオミキサを通じてミックスされ，イヤホンから聞こえる．

4.2　PC アプリケーション

PC アプリケーションの構成を図 5 に示す．アプリケーションは Python と openFrameworks を用いて作成した．アプリケーションは，PC1 で実行されるプレイヤー 1，PC2 で実行されるプレイヤー 2，PC3 で実行されるコントローラで構成される．プレイヤー 1 とプレイヤー 2 では音声・映像の録音・録画と再生が行われ，コントローラで再生の制御を行う．アプリケーションの処理フローを図 6 に示す．まず，プレイヤーで Web 会議映像の録音・録画を行い，音声データ・画像データとして保存する．音声データの再生速度を音程を変えずに上げるため，ソフトウェアライブラリ Rubber Band Library [20] を用いて音声処理を行う．また，音声認識ツールキット Vosk [21] を用いて，会議音声を文字起こしし，テキスト化する．2.0 倍速の音声データ，画像データ，テキストデータを再生処理部分で読み取り，プレイヤーで再生・表示する．複数の Web 会議への同時参加を支援する機能として，準リアルタイム視聴機能，タイムシフト視聴機能，リアルタイム視聴機能を実装した．機能の切り替えは，コントローラから任意のタイミングで行うことができる．

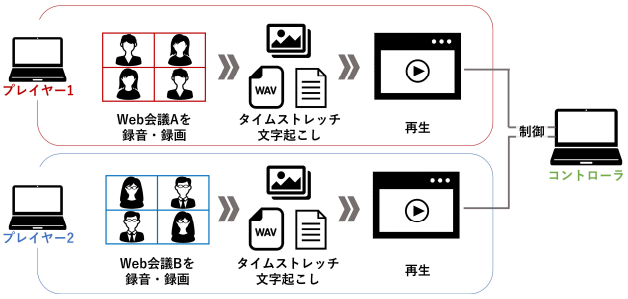

図 6　処理フロー

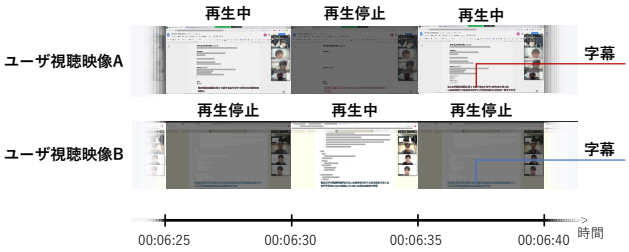

図 7　準リアルタイム視聴機能の再生イメージ

準リアルタイム視聴機能

準リアルタイム視聴機能の再生イメージを図 7 に示す．準リアルタイム視聴機能では，プレイヤー 1 とプレイヤー 2 で 2.0 倍速の会議映像が再生される．画面下部には会議の発言内容が字幕として表示される．2.0 倍速の会議映像はプレイヤー 1 とプレイヤー 2 でそれぞれ交互に再生される．交互再生の制御はコントローラで行う．

タイムシフト視聴機能

タイムシフト視聴機能では，Web 会議映像 A または Web 会議映像 B いずれか一方を早戻しし，視聴できる．この機能により，会議中聞き逃してしまった内容を何度も視聴できる．このとき，再生速度は 1.0 倍速，1.5 倍速，2.0 倍速の中から任意で選択できる．

リアルタイム視聴機能

リアルタイム視聴機能では，Web 会議映像 A または Web 会議映像 B いずれか一方のリアルタイムの映像を視聴できる．この機能により，リアルタイムの Web 会議へ発言や返答ができる．また，マイクミュートの切り替えをコントローラで制御する．

5.　評価実験

第一著者をシステムユーザとして，図 8 のように実装したシステムを用いて同時に行われる二つの Web 会議へ同時に参加できるか評価する実験を行った．被験者には，システムユーザが違和感なく二つの Web 会議に同時に参加できていたか，評価してもらった．

5.1　実験内容

被験者は，20 代の男性 6 名である．6 名の被験者をそれぞれ 3 人ずつグループ A とグループ B に分け，議題に沿っ

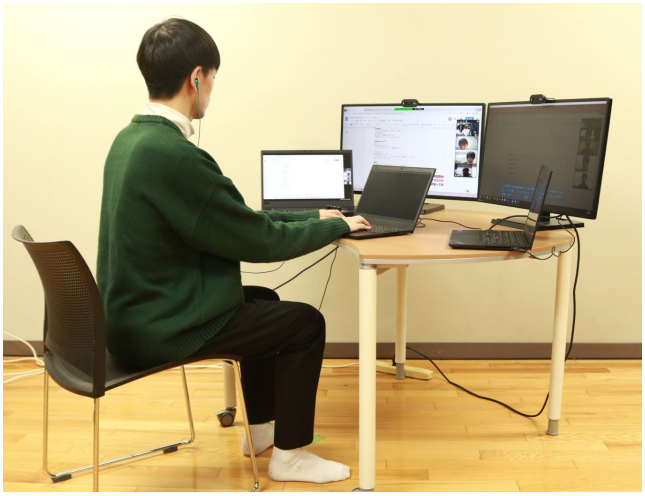

図 8 システムを使用している様子

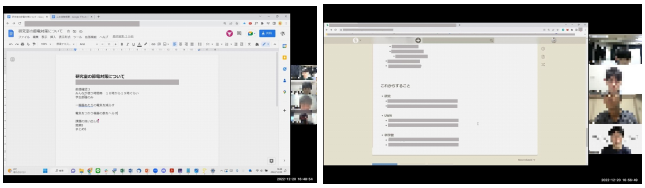

図 9 Web 会議の様子

て会議をしてもらった．グループ A とグループ B の Web 会議の様子を図 9 に示す．また，グループ A の議題は「研究室の節電対策のために何をするべきか」，グループ B の議題は「研究の進捗報告」とした．会議の時間は 20 分間とし，会議終了後，システムユーザは違和感なく会議に参加できていたか，システムユーザに対して思ったことを答えてもらうインタビューを行った．

5.2 実験結果と考察

実験結果から考察を行い，システムに必要な機能について議論する．インタビューから，「システムユーザの反応が遅く会議が止まることがあった」という意見があった．また，「システムユーザの名前を読んでから返事が返ってくるまでに 20 秒間の程度の遅れが生じていた」という意見があった．提案システムでは，準リアルタイム視聴機能を用いて，会議映像を視聴し，倍速再生された映像内で名前を呼ばれたり，意見を求められたときに，リアルタイム視聴機能を用いてリアルタイムの会議に発言を行うことを想定している．しかし，システムユーザ以外の会議参加者が違和感を覚える程度の遅延が発生していることから，リアルタイムの会議映像で返事や意見を求められたときにすぐにリアルタイムの会議へ発言ができる機能が必要である．

6. まとめ

本論文では，複数の Web 会議に同時に参加している状況において，参加しているすべての Web 会議の内容を理解し，発言などのインタラクションを支援するシステムを提案した．提案システムを実装し，複数の Web 会議に同時に参加している状況において，システムが有効であるか第一著者をシステムユーザとして評価実験を行った．実験の結果，提案システムを用いてユーザが複数の Web 会議内容の理解することを支援できたが，一方，リアルタイムの会議への発言や応答に 20 秒程度の遅れが生じ，ユーザ以外の会議参加者に違和感を与える結果になった．今後はシステムの機能を改善・追加し，ユーザ以外の会議参加者から返事や意見を求められたときにすぐに反応し，発言を支援するシステムを構築する．また，ブレーンストーミング，報告や連絡を行う会議，意思決定を行う会議など，どの会議の形式が同時参加に適しているか調査する．

謝辞 本研究の一部は，JST CREST(JPMJCR18A3) の支援によるものである．ここに記して謝意を表す．

参考文献

[1] 東京都福祉保健局: 新型コロナウイルス感染症について, 入手先 ⟨https://www.fukushihoken.metro.tokyo.lg.jp/iryo/kansen/corona$_p$ortal/info/shingatakorona.html⟩ (Accessed 2022-12-20).
[2] 高田 格, 栖関邦明, 杉山阿葵, 岡田謙一: 2 つの遠隔会議への同時参加支援手法, 情報処理学会論文誌, Vol. 50, No. 1, pp. 236–245 (Jan. 2009).
[3] 山本貴文, 土田修平, 寺田努, 塚本昌彦: 複数の Web 会議へ同時参加するための高速再生・映像切替方式, マルチメディア，分散協調とモバイルシンポジウム 2021 論文集, Vol. 2021, No. 1, pp. 140–148 (June 2021).
[4] 長濱 澄, 森田裕介: 映像コンテンツの高速提示による学習効果の分析, 日本教育工学会論文誌, Vol. 40, No. 4, pp. 291–300 (Mar. 2017).
[5] 栗原一貴, CinemaGazer: 動画の極限的な高速鑑賞のためのシステムの開発と評価, コンピュータソフトウェア, Vol. 29, No. 4, pp. 293–304 (Nov. 2012).
[6] 粥川青汰, 樋口啓太, 米谷 竜, 中村優文, 佐藤洋一, 森島繁生: 物体検出とユーザ入力に基づく一人称視点映像の高速閲覧手法, 研究報告コンピュータビジョンとイメージメディア (CVIM), Vol. 2017, No. 4, pp. 1–8 (Nov. 2017).
[7] Y. J. Lee, J. Ghosh, and K. Grauman: Discovering Important People and Objects for Egocentric Video Summarization, *Proc. of the Conference on Computer Vision and Pattern Recognition*, pp. 1346–1353 (June 2012).
[8] 中野裕太, 服部 哲, 速水治夫: ゲーム実況動画における動画多画面視聴支援システムの提案, 分散協調とモバイルシンポジウム 2011 論文集, pp. 370–373 (June 2011).
[9] P. Y. Hsueh and J. D. Moore: Improving Meeting Summarization by Focusing on User Needs: A Task-Oriented Evaluation, *Proc. of the 14th International Conference on Intelligent User Interfaces*, pp. 17–26 (Feb. 2009).
[10] 水田賢志, 菱山玲子: 電子会議への途中参加支援のためのダイジェスト提示システムの効果, 人工知能学会全国大会論文集, pp. 1–4 (June 2011).
[11] S. Tuker, O. Bergman, A. Ramamoorthy, and S. Whittaker: Catchup: A Useful Application of Time-Travel in Meetings, *Proc. of the 2010 ACM Conference on Computer Supported Cooperative Work*, pp. 99–102 (Feb.

2010).
[12] Y. Shi, C. Bryan, S. Bhamidipati, Y. Zhao, Y. Zhang, and K. L. Ma: MeetingVis: Visual Narratives to Assist in Recalling Meeting Context and Content, *Journal of IEEE Transactions on Visualization and Computer Graphics*, Vol. 24, No. 6, pp. 1918–1929 (June 2018).
[13] A. E. James, A. G. Nanos, and P. Thompson: V-ROOM: A Virtual Meeting System with Intelligent Structured Summarisation, *Journal of Enterprise Information Systems*, Vol. 10, No. 8, pp. 863–892 (Oct. 2016).
[14] 安西 悠, 江木啓訓, 西川真由佳, 湯澤秀人, 松永義文, 岡田謙一: 遠隔会議への同時多重参加に関する基礎検討, 情報処理学会研究報告グループウェアとネットワークサービス (GN), Vol. 2005, No. 30, pp. 75–80 (Mar. 2005).
[15] 高田 格, 杉山阿葵, 岡田謙一: 変速再生と映像切替による多重会議支援手法の提案, 情報処理学会研究報告グループウェアとネットワークサービス (GN), Vol. 2007, No. 56, pp. 67–72 (June 2007).
[16] Zoom Video Communications: One platform to connect, 入手先 ⟨https://zoom.us/⟩ (Accessed 2022-12-20).
[17] Microsoft Teams: リモート ワーク - コラボレーション ツール, 入手先 ⟨https://www.microsoft.com/ja-jp/microsoft-teams/group-chat-software⟩ (Accessed 2022-12-20).
[18] E. Foulke and T. G. Sticht: Review of Research on the Intelligibility and Comprehension of Accelerated Speech, *Journal of Psychological Bulletin*, Vol. 72, No. 1, pp. 50–62 (July 1969).
[19] J. L. Flanagan and R. M. Golden: Phase Vocoder, *Journal of Bell System Technical Journal*, Vol. 45, Issue. 9, pp 1493–1509 (Nov. 1966).
[20] Breakfast Quay: Rubber Band Library, 入手先 ⟨https://breakfastquay.com/rubberband/⟩ (Accessed 2022-12-20).
[21] Alpha Cephei: Vosk, 入手先 ⟨https://alphacephei.com/vosk/⟩ (Accessed 2022-12-20).