「マルチメディア，分散，協調とモバイル
(DICOMO2021) シンポジウム」 令和3年6月-7月

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

山本貴文[1]　土田修平[1]　寺田 努[1]　塚本昌彦[1]

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を目的とする．本稿では，会議での発言において重要となる会議内容の理解に着目し，リアルタイム性を維持した会議内容の理解を支援するシステムを提案する．提案システムは録画された二つの会議映像の音声データを分析し，音声の有無から必要な場面のみを抽出することで，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．システム設計に向けて，音声情報を含むシーンの抽出，再生速度，映像切替，字幕の各条件が，二つの Web 会議の内容理解に与える影響を調査した．評価実験において，被験者は二つの会議映像を視聴した後，会議内容が理解できたかを確認する問題に解答した．実験結果から，音声情報を含む部分のみを抽出した 2.0 倍速の会議音声を交互に再生しながら会議の字幕を提示することで，一つの会議の視聴時間で二つの会議内容を 8 割程度理解できることがわかった．

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リモートワークが推進され，Web 会議システムの利用機会が増えている．Web 会議システムはインターネットを通じて，音声・映像のやり取りや資料共有を行うコミュニケーションツールである．米マイクロソフトによると，Web 会議システム「Microsoft Teams」の 1 日当たりの利用者数が 2020 年 3 月 11 日からの 1 週間で 1200 万人増加し，Web 会議の普及率は高まっている [1]．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持つ．そのため，従来の対面による会議と比べ，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．さらに，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 [2] を防ぐことができる．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．

会議は企業や大学などの組織において重要なプロセスであり，意思決定や情報伝達のための有効な手段であるが，参加する必要のある会議にすべて参加できるとは限らない．例えば，ダブルブッキングや予定の時刻より会議の時間が長引くことにより，二つの異なる会議が重なってしまうという問題がある．ここで Web 会議であれば，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できるため，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 [3] は，2.2 倍速の二つの遠隔会議映像を 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，操作性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 会議ツールに無いため，複数のツールを使用しなければならない．発言を行う場合，対象の会議のマイクをオンにし，その他の会議のマイクはオフにして発言が聞こえないようにする必要がある．そのような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

[1] 神戸大学大学院工学研究科
Graduate School of Engineering, Kobe University

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，筆者らの知る限りあまり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，会議への発言において重要となる会議内容の理解に着目し，リアルタイム性を維持した会議内容の理解を支援するシステムを提案する．提案システムは，録画された二つの会議映像の音声データを分析し，音声の有無から必要な場面のみを抽出することで，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．

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．綿森ら [4] は，健常者 20 名，失語症患者 20 名を被験者として，話す速度を変化させた時の理解度を調査した結果，話す速度を 50 %圧縮しても，健常者は理解力に負荷がかからなかったと報告している．同様に再生速度が内容理解に与える影響を調査した研究がある．長濱ら [5] 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 [6] 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 [7] 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 [8] 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 [9] 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 [10] 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 [11] 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 [12] 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 [13] 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 [14] 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．Girgensohn ら [15] は，過去に開催された複数のビデオ会議映像の間にハイパーリンクを自動生成する HyperMeeting を開発し，過去の会議から取得したい映像情報へのアクセスを容易にした．

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，会議映像の音声の有無から必要な場面を抽出し，再生速度を上げることで，会議内容の把握にかかる時間の短縮を試みる．

2.3 会議音声から参加者の行動や感情をフィードバックする研究

会議の音声データから会議参加者の行動や感情を分析し，そこから得た情報をユーザにフィードバックする研究は数多く行われている．Cowell ら [16] は，言語処理ツール ChAT を用いて会議音声の分析を行い，会議中の参加者の立場や感情を表示した．これは，参加者の発言が他者の感情に与える影響や発言内容が質問であるか，否定的であるかを示すことで，会話の流れやコミュニケーションを視覚化した．会議参加者の発言内容や感情に着目したフィードバックの研究として，Samrose ら [17] は，ビデオ会議中に得られた情報を会議終了後にフィードバックすることで，次の会議での行動変容を促した．これは，全会議時間のうち発言した時間の割合，発言に対して割り込んだ，または割り込まれた回数，発言者の移り変わり，怒りや驚きのような感情，笑顔の回数の五つを参加者に提示することで，積極的な発言や他者に発言を促すといった行動変容がフィードバックの前後で観察された．会議での行動変容の別アプローチとして，Bergstrom ら [18] は，会議中の 4 人の発言量をテーブル上に投影することにより，会議への貢献度を可視化する Conversation Clock を提案した．この研究では，貢献度の少ない被験者に対して発言を促すことにより，会議全体の発言量を均一化する手法が検討されている．他にも Aseniero ら [19] は，オンライン会議への積極的な関与を促進する Meet Cues を提案した．Meet Cues は，他者の発言に対して「いいね」や「わからない」といったリアクションを匿名で行う機能やコメント機能を実装し，会議への積極的な関与を促進した．また，1 分間あたりのリアクションの量をタイムラインで表示し，リアクションが多い部分を録音できる機能を搭載している．

しかし，これらの研究は発言量や表情のような態度しか評価されておらず，会議内容を理解できているかは考慮されていない．本研究では，会議の音声データを分析し，音声認識技術を用いて会議の字幕を提示することで，会議内容の理解を補助することを目指す．

2.4 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 [20] 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 [3, 21] 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．

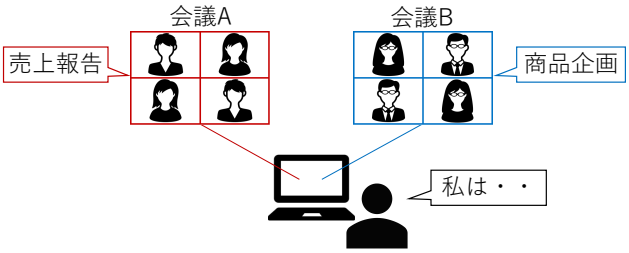


図 1: 想定環境

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. システム検討

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の構築に向けて，システムの想定環境，要件，及びその解決方法について整理する．

3.1 想定環境

想定環境を図 1 に示す．本研究では，1 台の PC を用いて二つの Web 会議に同時に参加する状況を想定している．Web 会議ツールはそれぞれの会議において，異なるツールを使用した．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．また，参加している Web 会議内のユーザの発言や返答は通常の会議と同様に行う．

3.2 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 [22] や Google Meet [23] のような Web 会議ツールを用いる場合，複数のウィンドウを同時に操作する必要がある．会議内容を把握しつつ，マイクの切り替えやウインドウの配置のような操作を行うことは難しい．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

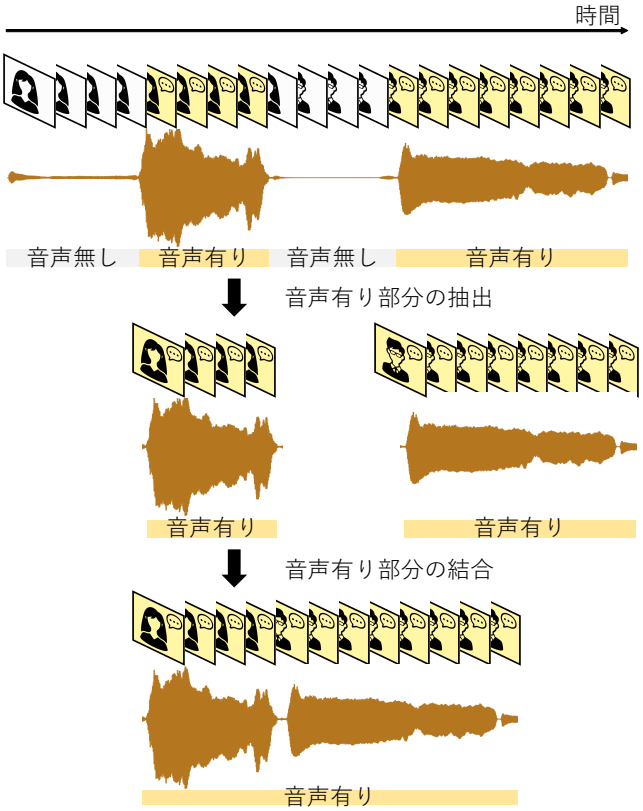


図 2: 音声有りの部分の抽出

3.3 解決方法

前節のシステム要件を踏まえて，システムのプロトタイプを提案する．まず，二つの会議映像の音声データを分析し，音声の有無から必要な場面のみを抽出する．Web 会議では，発言のタイミングを伺う時間や発言を考える時間により，誰も発言をしない，音声情報を含まない部分が現れる．そこで画面録画により会議映像をリアルタイムに取り込み，音声データの振幅から音声情報を含む部分と音声情報を含まない部分に分割する．そのうち音声情報を含む部分のみを抽出し，それぞれを結合した会議映像を作成する．音声情報を含むシーンの抽出の流れを図 2 に示す．

次に，抽出した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．図 3 に示すように，交互に視聴する場合，初めはリアルタイムから大きく遅れずに会議映像を視聴できるが，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，会議映像の再生速度を上げることで，映像視聴に必要な時間を短くする．再生速度は内容を聞き取れる範囲の再生速度とする．例えば再生速度を 2.0 倍速にすると，一つの映像を通常速度で視聴する時間で二つの会議映像を視聴できる．また，先行研究 [3] では，2.2 倍速の二つの遠隔会議映像を 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

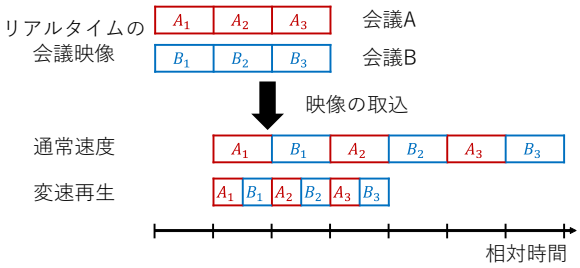


図 3: 倍速再生のイメージ

し，リアルタイムとの遅れをより短くすることを試みる．

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 [24]，PSOLA と呼ばれる音声圧縮技術 [25] を用いて，音程を変えずに再生速度を変える．これにより，内容理解を容易にする．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，異なる Web 会議ツールのマイクのオン・オフのスイッチを 1 画面に配置することで発言を行う操作を容易にする．

4. 評価実験

提案システムの構築に向けて，二つの会議映像を視聴する際の音声情報を含むシーンの抽出をし，再生速度，映像の切替間隔が内容理解に与える影響を明らかにするための実験を行った．表情や身振り手振りを含む視覚情報による効果を排除するために，音声のみの映像を用いた．字幕を付けた音声のみの映像の再生速度と切替間隔を変化させることで，内容理解を損なわない再生速度と映像切替間隔の組合せを調査する．

4.1 実験準備

再生速度と切替間隔を各 3 種類用意し，それぞれを組み合わせた 9 通りの映像を作成した．また，二つの会議映像を単純に繋げたものと二つの映像を同時に並べた映像を作成し，以下合計 11 種類の映像[*1]を用意した．

- 1.0 倍速の二つの映像を同時に視聴
- 1.0 倍速で片方の映像を視聴し終えた後，もう片方の映像を視聴
- 1.0 倍速の二つの映像を 3 秒間隔で交互に視聴
- 1.0 倍速の二つの映像を 5 秒間隔で交互に視聴
- 1.0 倍速の二つの映像を 10 秒間隔で交互に視聴
- 1.5 倍速の二つの映像を 3 秒間隔で交互に視聴
- 1.5 倍速の二つの映像を 5 秒間隔で交互に視聴

[*1] 作成した視聴映像，
https://drive.google.com/drive/folders/1OvLfYNicBeTUD5kmMWHyljMuBUZ8UsEA?usp=sharing

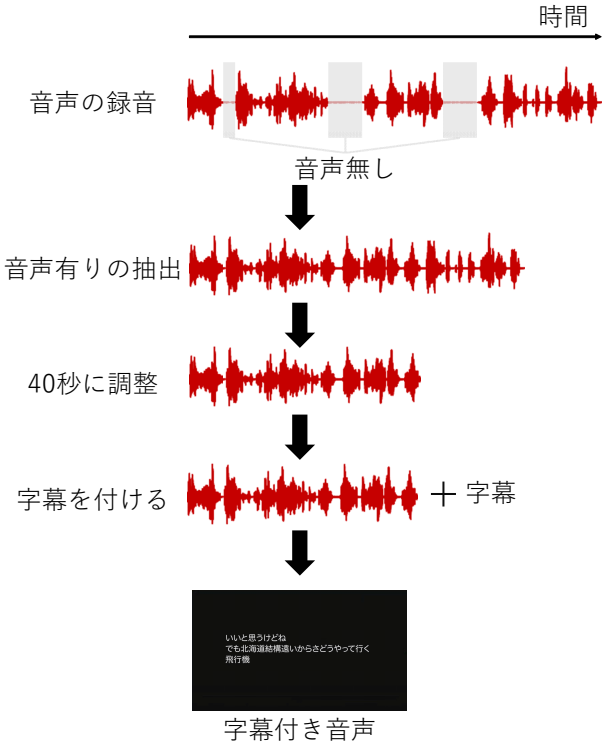


図 4: 字幕付き映像の作成

- 1.5 倍速の二つの映像を 10 秒間隔で交互に視聴
- 2.0 倍速の二つの映像を 3 秒間隔で交互に視聴
- 2.0 倍速の二つの映像を 5 秒間隔で交互に視聴
- 2.0 倍速の二つの映像を 10 秒間隔で交互に視聴

実験で使用した映像の準備手順を以下に示す．まず Zoom を用いて会議映像を録画した．他の Web 会議システムツールと比較して，映像の録画が容易であるため，本実験では Zoom を使用した．そして，動画編集プログラム Vrew [26]，映像編集ソフト Adobe Premiere Pro [27]，マルチメディア編集ソフト FFmpeg [28] を用いて，動画の編集，字幕の付加を行った．

4.1.1 音声と字幕を付けた映像の作成方法

図 4 に字幕付き映像作成の流れを示す．まず，Zoom を用いて，あるトピックに沿った会議音声の録音を 60 秒間行った．次に，Vrew を用いて録音された音声から音声情報を含まない部分をカットした．このとき，音声の長さは 40 秒程度になった．そして，Adobe Premiere Pro を用いて音声に字幕を付け，字幕付き映像を作成した．会議内容のトピックは慣れによる内容理解の差を低減するため 6 パターン用意した．会議に用いたトピック A〜F の 6 種類を表 1 に示す．A〜C は複数人がブレーンストーミングを行うようなトピック，D〜F は 1 人が報告を行うようなトピックに設定した．

4.1.2 視聴映像の作成

図 5 に視聴映像作成の流れを示す．前小節で作成した字幕付き映像の再生速度を 1.0 倍速，1.5 倍速，2.0 倍速にした 3 種類の映像を作成した．次に再生速度を変化させた映像を t 秒間隔（t=3, 5, 10）で分割した．次に分割した映像それぞれの最後に t 秒間の音声無しの静止画を挿入した．静止画を挿入した理由は，それらすべての映像を結合し，音声有りの映像と音声無しの静止画が t 秒間隔で交互に繰り返される映像を作成するためである．映像は再生速度と t の値を変えて 1 トピックにつき 9 パターン作成した．最後に二つのトピックの会議音声が交互に再生されるように視聴映像を組み合わせた．二つのトピックの組合せは表 1 より A と B，C と D，E と F とした．完成した視聴映像の再生イメージの例を図 6 に示す．例えば，同時視聴では二つの 60 秒の映像を同時に視聴する．11 種類の視聴方法のうち同時視聴を，二つの映像を単純に同時に視聴する従来手法として扱った．このとき，提案手法と比較を行うため，音声情報無しの部分をカットしなかった．再生速度が 1.0 倍速，映像切替間隔が 40 秒の場合，1.0 倍速の会議映像 A を 40 秒視聴してから 1.0 倍速の会議映像 B を 40 秒視聴する．再生速度が 1.5 倍速，映像切替間隔が 5 秒の場合，1.5 倍速の会議映像を，A，B，A，と 5 秒ずつ視聴する．

表 1: トピック

トピック	内容
A	旅行の計画
B	掃除の日程
C	イベントの企画
D	自己 PR
E	研究報告
F	アルバイトの相談

4.2 実験方法

被験者は，20 代の男性 10 名と女性 1 名の合計 11 名である．そのうち 5 名は対面，6 名は Zoom を用いてオンラインで実験を行った．実験風景を図 7 に示す．まず，被験者に PC の前に座ってもらい，実験に関する簡単な説明と別途用意したサンプル動画で PC の音量を内容が理解できる程度に調整してもらった．そして，前節の手順で用意した 11 種の映像のうち，各被験者につきランダムで選択した 3 種の映像を視聴した．視聴中，被験者が内容に関する記録を取ること，早送り，巻き戻し，一時停止のような再生操作を行うことを禁止した．各映像の視聴を終える毎に，被験者は内容についての理解を確かめる問題を解答した．対面での被験者は紙に解答し，オンラインでの被験者は Google Form に解答した．トピック A〜F に関する具体的な設問内容を表 2 に示す．また，説明の際に，映像は音声情報を含まない部分をカットし，字幕をくわえていること，再生速度と音声の切り替え間隔を調整していることを伝えた．また被験者が会議のおおまかな内容を把握するため，会議のトピックを映像視聴前に伝えた．設問内容に関する情報

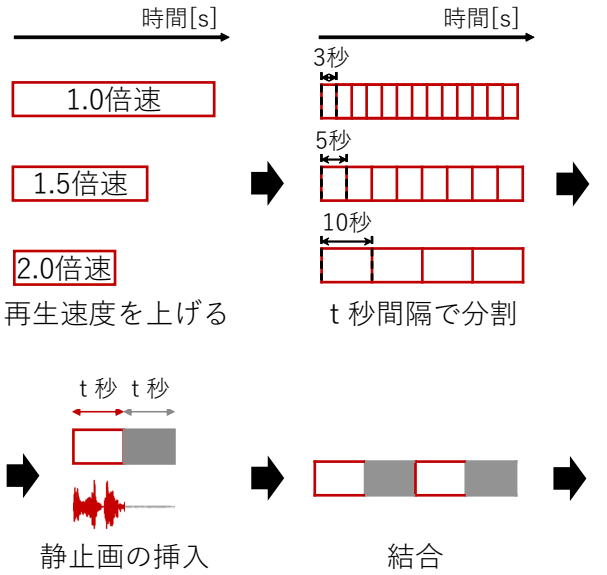


図 5: 視聴映像の作成

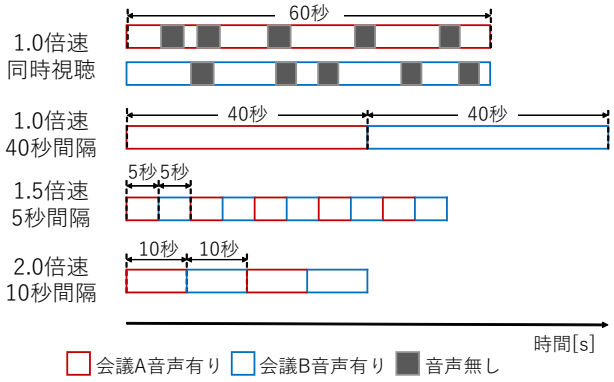
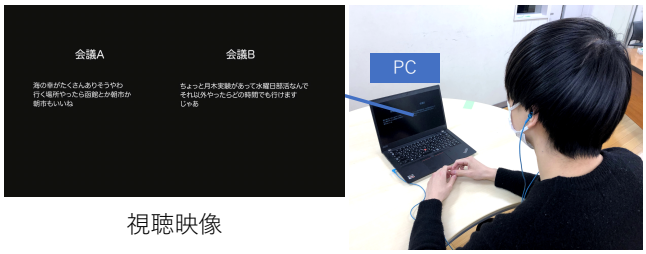


図 6: 視聴映像の再生イメージの例

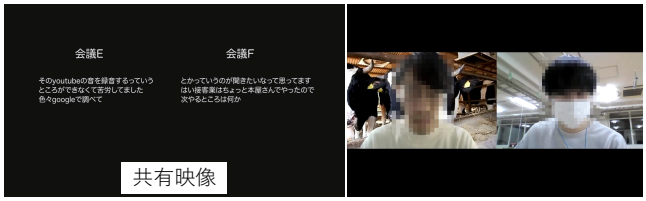
は伝えなかった．各映像 3 人の被験者による解答を得た．

4.3 実験結果と考察

図 8 に各再生速度と各映像切替間隔における設問の正解率を示す．また，各再生速度における映像 AB の総再生時間は図 9 のようになった．まず 1.0 倍速の結果に着目すると，映像切替間隔が 5 秒以上の場合，9 割以上と高い正解率を得られ，特に 5 秒の場合は会議内容を 100%理解できた．しかしこの場合では，二つの映像を視聴する時間が必要となることが問題点として挙げられる．また，3 秒間隔で交互に視聴する場合では，同時視聴よりも正解率が低



(a) 対面



(b) オンライン

図 7: 実験の様子

表 2: 設問内容

音声	No.	設問項目
A	1	何月に旅行に行くと言っていたか
	2	どこに行きたいと言っていたか
	3	何をたくさん食べたいと言っていたか
	4	何の像を見に行くと言っていたか
	5	どうやって行くと言っていたか
B	1	どこの掃除をするか
	2	K 君の月曜日と木曜日の予定は何か
	3	何曜日に掃除をすることになったか
	4	何時からすることになったか
	5	K 君は何をすると言っていたか
C	1	会議で出た企画を 5 個全て答えよ
D	1	彼の長所は何と言っていたか
	2	もう一つの長所は何か
	3	学部時代，彼が取り組んでいた研究は何か
	4	現在，彼が取り組んでいる研究テーマは何か
	5	現在，彼は何の開発をしているか
E	1	彼は何の音を録音しようとしていたか
	2	何を用いて録音しようとしていたか
	3	何の音は録音できたと言っていたか
	4	何で調べたと言っていたか
	5	これから何をすると言っていたか
F	1	現在，彼は何のアルバイトをしているか
	2	なぜアルバイトを辞めたいと言っていたか
	3	先輩に何を尋ねたいと言っていたか
	4	なぜ次は接客業をしたくないと言っていたか
	5	次は何のアルバイトをしたいと言っていたか

くなった．これは音声情報の有る部分のみを抽出した映像を短い間隔で交互に視聴したことにより，情報を受け取り整理するための時間が少なく，内容理解が困難となったためであると考える．1.0 倍速の場合，切替間隔が 10 秒よりも 5 秒の方が高い正解率を得られた．一方，1.5 倍速の場合，切替間隔が 5 秒よりも 10 秒の方が高い正解率を得ら

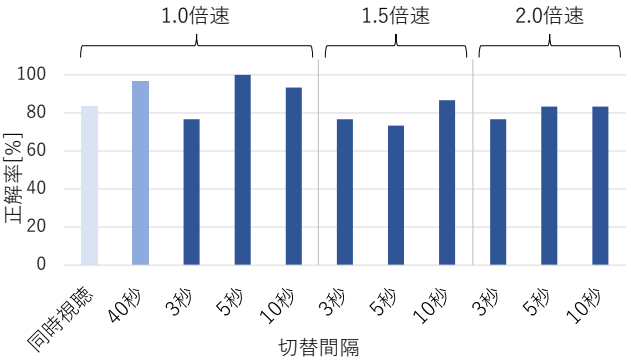


図 8: 各再生速度と各映像切替間隔における正解率

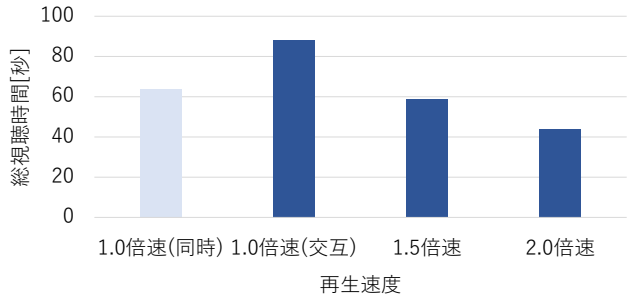


図 9: 各再生速度における映像の総再生時間

れた．また，切替間隔が 5 秒の場合，1.0 倍速，2.0 倍速，1.5 倍速の順に正解率が高かった．このように，再生速度と正解率，切替間隔の長さと正解率に直接的な関係は見られなかった．これは，会話の切れ目の違いが内容理解に影響を与えたため，内容理解のための適切な切替間隔の長さは変化すると考える．

次に 1.5 倍速の結果と 2.0 倍速の結果について比較を行う．切替間隔が 3 秒，10 秒の場合は同様の正解率となり，5 秒の場合は 2.0 倍速の方が正解率が高くなった．また「字幕を読むことで内容を理解しやすかった」，「音声を聞くことよりも字幕を読むことで内容を把握できた」という意見があった．音声の再生速度が上がると，聴覚で内容を理解することは困難であるため，内容理解における字幕の役割が 1.0 倍速に比べてより大きくなることが示唆された．よって，字幕を読む時間が確保できれば，二つの会議内容を同時に理解できると考える．

5. 今後の展望

4 章における実験結果と考察より，本研究で得られた結果の応用例と，今後の課題について検討する．

5.1 応用例

実験結果より，音声情報を含む部分のみを抽出した 2.0 倍速の会議音声を交互に再生し，会議の字幕を提示することで，一つの会議映像の視聴時間で二つの会議内容を理解できることを確認した．また字幕を提示することで二つの音声を同時に理解できることがわかった．今後は，二つの Web 会議に同時に参加する状況において，会議音声データから字幕を作成し，視聴者に提示することで，異なる二つの会議内容を同時に理解できると考える．また音声情報を含む部分のみを抽出することにより，会議映像を視聴する時間を減らし，再生速度を上げることなく，内容の適切な理解や発言に時間を充てることができると考える．さらに，複数の Web 会議に同時に参加することができれば，二つの異なる会議が重なった場合でもどちらかを欠席する必要がなくなる．それに加えて，二つの会議を同時に開催することで，会議にかける時間を減らすことができる．

5.2 課題

今後の課題点として，円滑なインタラクションを行うため，どのくらい返答が遅れると相手は違和感を感じるかを調査することが挙げられる．また，ユーザが片方の会議に対して発言を行う場合に，発言をしていない会議の内容を適切に理解するための手法を検討する必要がある．今回の実験では，映像の切替間隔を 3 秒，5 秒，10 秒とした．しかし，実際の Web 会議で交互に会議映像を視聴し，会議中に返答を求められた場合，映像の取込に要する時間や交互に視聴することにより，返答するまでに数秒間から十数秒間の遅れが生じる．それにより，会議参加者が返答が無いことに違和感が生じ，また会議の進行を妨げてしまう可能性がある．よって，映像を切り替える間隔を短くし，リアルタイムとの遅れを短くする必要がある．また，参加者が違和感なく，会議の進行を妨げない場合であれば，切替間隔を長くすることで，より会議内容の理解を容易にできると考える．

今回の実験では，予め録音・編集した映像を使用することで，被験者は映像の視聴に集中できたため，内容理解を大きく損なわない結果になったと考えられる．しかし，ユーザが片方の会議に対して発言を行う場合，発言の間は映像を視聴できず，情報が抜け落ちてしまう恐れがある．よって，交互に視聴する周期性を保ったまま，内容理解の補助を行う必要がある．例えば，会議映像の視聴できない部分の内容は要約を行い，テキストで表示することで，視聴できない部分の内容を補完する．これによって会議への発言をするために十分な内容理解の補助ができるかどうか調査する必要がある．また会議に意見や発言を行うために会議映像をすべて視聴する必要があるか，必要最低限の要約だけで十分であるかを調査する必要がある．

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．またシステム構築のために音声情報を含むシー

ンの抽出，再生速度，映像の切替間隔，字幕が二つの Web 会議の内容理解に与える影響を調査する実験を行った．本稿では予め録画した Web 会議映像を音声と字幕のみに編集し，再生速度と映像切替間隔を変化させた．そして，編集した動画を被験者が視聴後，会議内容に関する設問に答えてもらうことで，内容についての理解度を調査した．その結果，音声情報を含む部分のみを抽出した 2.0 倍速の会議音声を交互に再生し，会議の字幕を提示することで，一つの会議の視聴時間で，二つの会議内容を理解できることがわかった．また提案手法において，再生速度を上げると内容理解が困難になり，切替間隔を長くすると容易になるという結果は得られなかった．この結果から，再生速度と問題の正解率，切替間隔の長さと設問の正解率にそれぞれ直接的な関係は見られず，会話の切れ目の違いが内容理解に影響を与える可能性を確認した．

今後の課題として，実験結果をもとにシステムを実装すること，円滑なインタラクションを取るために何秒まで遅延が許容されるか検討することが挙げられる．またユーザが会議に参加し，発言を行う場合に，会議内容の理解を補助するシステムも設計する必要がある．設計に向け，ユーザが会議内で意見や発言を行うために必要な情報を調査する．例えば，会議へのスムーズな発言をするために，会議全体の映像を視聴する必要があるのか，概要を把握できれば問題ないのか，などを検討する必要がある．これにより，同時に参加している二つの Web 会議へのインタラクションを可能とする情報提示の方法を探り，システムの開発を行う．

謝辞

本研究の一部は，JST CREST(JPMJCR16E1，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] Microsoft365: Remote work trend report: meetings https://www.microsoft.com/en-us/microsoft-365/blog/2020/04/09/remote-work-trend-report-meetings/ (Accessed 2021-4-20).

[2] 東京都福祉保健局: 新型コロナウイルス感染症について, https://www.fukushihoken.metro.tokyo.lg.jp/smph/iryo/kansen/shingatakorona.html (Accessed 2021-4-20).

[3] 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).

[4] 綿森淑子, 笹沼澄子: 話しことばの速度の変化が失語症患者の理解力に及ぼす影響について, 音声言語医学, Vol. 15, No. 2, pp. 31–36 (July 1974).

[5] 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).

[6] 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).

[7] 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).

[8] 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).

[9] 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).

[10] 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).

[11] 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).

[12] 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).

[13] 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).

[14] 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).

[15] A. Girgensohn, J. Marlow, F. Shipman, and L. Wilcox: HyperMeeting: Supporting Asynchronous Meetings with Hypervideo, *Proc. of the 23rd ACM International Conference on Multimedia*, pp. 611–620 (Oct. 2015).

[16] A. J. Cowell, M. L. Gregory, J. Bruce, J. Haack, D. Love, S. Rose, and A. H. Andrew: Understanding the Dynamics of Collaborative Multi-Party Discourse, *Journal of Information Visualization*, Vol. 5, No. 4, pp. 250–259 (Dec. 2006).

[17] S. Samrose, R. Zhao, J. White, V. Li, L. Nova, Y, Lu, M. R. Ali, and M. E. Hoque: CoCo: Collaboration Coach for Understanding Team Dynamics during Video Conferencing, *Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 1, No. 4, pp. 1–24 (Jan. 2018).

[18] T. Bergstrom and K. Karahalios: Seeing More: Visualizing Audio Cues, *Proc. of IFIP Conference on Human-Computer Interaction*, pp. 29–42 (Sep. 2007).

[19] B. A. Aseniero, M. Constantinides, S. Joglekar, K. Zhou, and D. Quercia: MeetCues: Supporting Online Meetings Experience, *Proc. of the IEEE Visualization Conference (VIS)* (Oct. 2020).

[20] 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).

[21] 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).

[22] Zoom Video Communications: Zoom, https://zoom.us/jp-jp/meetings.html (Accessed 2021-4-30).
[23] Google: Google Meet, https://apps.google.com/intl/ja/meet/ (Accessed 2021-4-30).
[24] 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).
[25] R. W. L. Kortekaas and A. G. Kohlrausch: Psychoacoustical Evaluation of the Pitch- Synchronous Overlap-and-Add Speech-Waveform Manipulation Technique Using Single- Format Stimuli, *Journal of the Acoustical Society of America*, Vol. 101, No. 4, pp. 2202–2213 (Apr. 1997).
[26] VoyagerX: Vrew, https://vrew.voyagerx.com/ja/ (Accessed 2021-4-30).
[27] Adobe: Adobe Premiere Pro, https://www.adobe.com/jp/products/premiere.html (Accessed 2021-4-30).
[28] FFmpeg: FFmpeg, http://www.ffmpeg.org/ (Accessed 2021-4-30).