

遠隔ミーティングの記録・再生を支援するシステムの研究

栗原 崇[†] 井上 直人^{†,*}
葛岡 英明[†] 鈴木 栄幸^{††}

ネットワーク技術の発達に伴い、リアルタイム実画像通信技術を用いた遠隔地間でのテレビ会議がより身近になってきている。また、コンピュータ技術によって、ビデオ映像や手書きメモをデジタルなデータとして簡易に保存できるようになった。しかし、膨大なデータの中から参照したい映像を見つけ出すのは容易ではない。著者らはこの問題を解決するために、記録映像の音声から韻律情報を抽出し、話題の切り替わる境界を提示するシステムを構築した。本稿では、システム構成を紹介し、システムの評価実験について述べる。

Study on a system which supports recording and replaying of remote meetings

TAKASHI KURIHARA,[†] NAOTO INOUE,^{†,*} HIDEAKI KUZUOKA[†]
and HIDEYUKI SUZUKI^{††}

Recent advancement of the network technology made tele-conferencing systems more common to us. Furthermore, computer technology allows us to record all the video images and digitized memos on to a hard disk very easily. On the other hand, it is not easy to find the specific part from the enormous quantity of data. In order to alleviate this problem, we propose to detect boundaries between topics and use them as the index of the recorded conversation. In this paper, system architecture and the preliminary experiment using the system are described.

1. はじめに

ネットワーク技術の発達に伴い、リアルタイム実画像通信技術を用いた遠隔地間でのテレビ会議がより身近になってきている。また、記録メディアの大容量化や、タブレット PC, PDA の普及により、遠隔会議参加者は自発的な行為として、会議内容の忘却とメモの紛失を防止するために、会議の様子や手書きメモをデジタルに保存すると予想される。

ところで、話し合いの様子をもれなく撮影した映像には、話し合いにおける全ての情報が含まれている。しかし、そのデータ量は膨大なものになり、ビデオのような連続映像から必要な部分を探し出すことは容易

ではない。仮に会議中に記録されたメモを会議内容のインデクスとして利用して、メモをクリックすることによって、それに関連した議論が行われている場面を適切に再生することができれば、短時間で会議内容を把握できると期待される。

そこで、本研究では、会議における話題が切り替わる境界を検出し、これを映像再生の候補点とすることによって、短時間で会議内容を把握できるシステムの開発を目的とする。

本稿では、まずはじめに研究の背景について述べ、先行研究における問題点を明らかにし、これを解決するためのシステムについて述べる。さらに予備実験を通じてシステム開発の展望について述べる。

2. 研究の背景

2.1 映像とメモを関連づけるシステムの研究

倉本らは会合情報記録検索システム ReSPoM¹⁾を開発した。図 1 に ReSPoM の概観を示す。ReSPoM では、議論の記録として音声のみを蓄積している。このシステムでは、紙上の環境と同じように資料やノートに対してメモを書き込むことができるとともに、会

[†] 筑波大学大学院システム情報工学研究科
Graduate School of Systems and Information Engineering,
University of Tsukuba

^{*} 現在、ソニー株式会社 ホームエレクトロニクスネットワークカンパニー

Presently with Sony Corporation Home Electronics
Network Company

^{††} 茨城大学人文学部
The Faculty of humanities at Ibaraki University

合参加者が明示的にメモや資料とそれに関連する発言との関連付け(リンクの作成)ができるようになっており、このリンクを用いることで、資料やメモを介して目的の発言を参照することができるようになっていいる。このような手法を採用することで、メモとそのメモのもとになる発言との確実な関連付けを行なうことができるが、関連付け作業に傾注する必要があるため発言がしづらかったという問題点が報告されている。

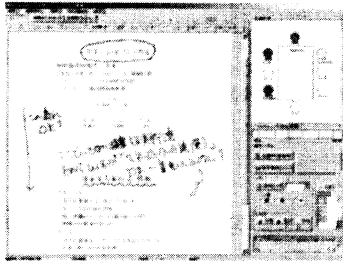


図 1 ReSPoM

Sifelman らは、実際のノートに書いた文字と音声とを関連付けることができる AudioNotebook²⁾を開発した。図 2 に AudioNotebook のシステムの概観を示す。このシステムでは、人間がメモを取るという行為に着目し、音声とメモとを関連付け、音声を蓄積する。これによってユーザはメモをインデックスとして、メモに関連付けられている音声を聞くことができる。AudioNotebook は、メモを取るという自然な行為をトリガとしているため、利用者に負担を掛けることなく過去に行なった会話を参照することができるという利点がある。

しかし、日常的な話し合いでは、作業者が指示対象物に対して指さしを行ないながら「この」や「あの」といった指示代名詞を用いて指示をしたり、物の大きさや作業の方向などを言葉とともにジェスチャを交えて説明したりすることが考えられる。このような場面においては、音声のみの提示では過去の作業内容を想起することが困難であると考えられる。そこで、本研究では、Audio Notebook の利点を生かしつつ、音声と映像の両方を用いることで、日常的な話し合いにおける議論の記録・再生を支援する。

2.2 同時発話を考慮した再生システムの研究

本研究室では、井上らによって、同時発話を考慮した遠隔ミーティングの記録・再生システム³⁾が開発されている。このシステムでは、会議中の映像とタブレットディスプレイへ書き込んだ手書きメモの関連付けを行い、メモから映像を参照できるようになっている。そして長時間の会議の中からメモの論拠にあたる

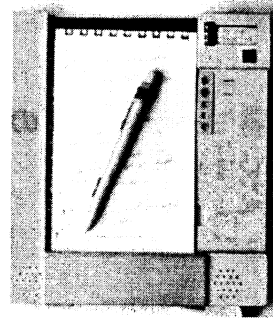


図 2 AudioNotebook

映像を探索するために、同時発話が頻繁に起きている場面を議論の中心と捉え、メモから映像を参照する際の候補点として提示している。この再生提示システムによる実験結果から、話し合いの中心場面を参照できるという利点が確認された反面、いきなり話し合いの中心となる場面が提示されるので内容理解のためには少し巻き戻す必要があるとの見解が得られている。

本研究では、井上らと同様にインフォーマルな会議中のメモが電子化される環境を想定し、会議中にユーザが ReSPoM のような明示的な操作を行なう必要がなく、再生時にメモと関連性の高い場面の映像を提示することで、会議終了後の再生支援を行なうことを目的とする。特に本研究では、井上らの研究で挙げられた、話し合いの中心場面で再生されるので内容理解のためには巻戻さなくてはならないという問題点を解消するために、図 3 のように、話題が転換している時刻を再生開始候補点として提示することでメモに関連深い話題の冒頭部分から映像を再生できるシステムを構築する。

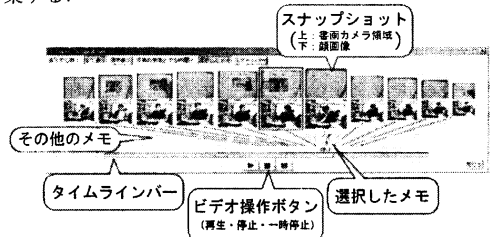


図 3 再生開始候補点提示ウィンドウ

3. 談話構造と韻律情報の関係

過去の研究において、日本語対話における談話と韻律情報との関係が示されている。小磯・下嶋・片桐は日本語対話中の情報境界と発話速度との関係を調べ、新たな情報単位の境界では減速し、情報境界でないならば加速するという関係が高確率で起こるとし、複数の話者による協調的調整によっても成立するとしてい

る⁴⁾。また菊池らは、二名の話者間で行われた課題遂行対話コーパスの五つの対話を分析し、発話の終了部分のパワーや基本周波数、ならびに談話標識を除いた発話の開始部分のパワーや基本周波数と、話題の切れ目の間に相関があることが示されている。また、話者間のポーズ長と話題の切れ目の深さにも相関があることが示されている⁵⁾。

本論文では、これまでに示されている韻律の特徴の中でも、発話開始部分の音声パワー、発話終了部分の音声パワー、発話ポーズ長の3つの要素から話題の境界を推定するアルゴリズムを開発し、映像の再生開始候補点に利用する。

4. システムの概要

本研究では、遠隔会議システム Agora⁶⁾の機能を拡張することで、インフォーマルな話し合いに適したシステムの構築を目指す。ここでは、システムの概要について述べる。

4.1 遠隔会議システム Agora

遠隔会議システム Agora(図4)は、遠隔作業用投影スクリーン、半共有領域、書画カメラ領域の3つの部分から構成されている。このシステムは、遠隔地にい



図4 遠隔会議システム Agora

る共同作業を等身大で投影することが可能な正面のスクリーンに加え、共同作業者が実物体や文書を映像として共有しつつ相互に指さしが可能な書画カメラ領域と、詳しい作業内容までは分からないが、なんとなく作業内容を把握することのできる半共有領域を備えている。これらの領域は、正面のスクリーンに投影された遠隔地の共同作業者の画像との身体的な位置関係を考慮して配置されているため、視覚的な臨場感だけではなく、作業者らの興味に応じて変化する身体の向きを正確に表現することも可能となっている。

さらに井上らは Agora に、会議中のメモと、会議中正面に投影される映像や書画カメラ領域の映像を関連付けるシステムを構築した⁷⁾。図5にこのシステムの概要を示す。

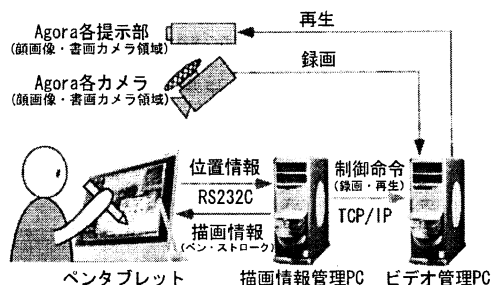


図5 映像とメモを関連づけるシステム

このシステムでは、作業者がメモを書くという行為をトリガとして、メモと録画した映像との関連付けを行ない、映像をビデオサーバに蓄積する。蓄積された映像は、ユーザがメモをペンでタップすることで、タップしたメモに関連付けられている時刻の映像が再生される。また、本研究ではこのシステムで記録される映像から音声抽出して操作することにより、発話音声の韻律情報を検出する。

4.2 再生開始時刻決定手法の概要

本節では、本研究が提案する再生開始時刻決定手法について述べる。図6に再生開始時刻決定手法の概要を示す。この手法では、再生開始時刻の候補を以下の手順で決定する。

- (1) 音声データを PC に取り込む。
- (2) 音声データを用いて、各作業者が発話中か否かの判定を行なう。
- (3) 音声データと発話状況から、発話単位のみとまりを作成する。
- (4) 発話単位に基づいて開始音量、終了音量、発話ポーズ長を算出する。
- (5) 4. で算出された各パラメータから、話題境界の候補を推定し、トピック提示ウィンドウに提示する。

話題境界の候補を推定するための音声データは、44.1kHzのサンプリングレート、16bit データでサウンドカード経由で取得している。さらに、取得した音声データの 100msec 毎の音量の平均値から、閾値によって話者が発話状態か無音状態かを判定し、それによって話者情報を管理するとともに、無音時間長を算出する。また、取得した 100msec 毎の音量の最大値と話者情報から、各発話単位の発話開始部分と終了部分の最大音量を算出する。こうして算出された発話単位ごとのデータをもとに、話題境界の候補を推定し、再生開始時刻の候補点とする。

5. 予備実験

本システムにおける再生開始候補時刻の決定手法が

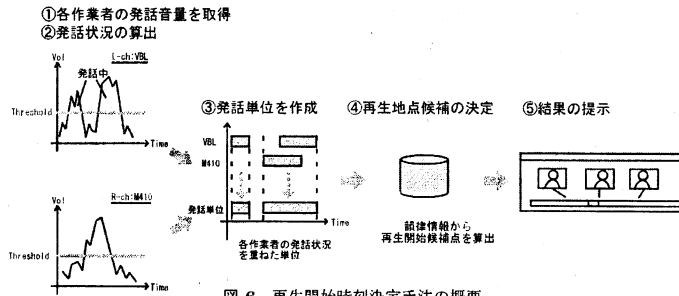


図6 再生開始時刻決定手法の概要

妥当なものであるかを調べるため、以下の2つの予備実験を行なった。本章では、本システムの利点、および問題点について議論し、今後の展望について述べる。

5.1 話題境界に関する調査実験の概要と結果

話題境界検出アルゴリズムの妥当性を確認するために、20分の会議映像から話題が大きく転換している箇所を調査する実験を行い、本アルゴリズムが検出した候補点と比較し評価した。

実験では、約20分間の遠隔会議の映像を、被験者に冒頭から末尾まで見てもらう。その際、被験者は話題が変わったと感じた時刻とその後どのような内容に変わったかをレポートに記述する。

被験者にはレポート作成終了後インタビューを行い、このレポートを基に会議において最も話題が大きく変わった時刻を5個指定してもらう。再生映像は2種類用意し、被験者には1つの映像が終了する毎にインタビューに答えてもらう。

なお作業について、話題が変わった時刻は会話の開始時刻を正確にとってもらうよう指示する。

筑波大学工学システム学類の学生を中心に8名(男女各4名)の被験者に対して実験を行なった。再生映像は2種類用意し、4名(男女各2名)ずつ見せる映像の順序を変えた。

本実験において作成されたレポートには話題の転換した時刻が記されており、これについては実験前に会話の開始時刻を正確にとるよう指令した。しかし、明らかに同じ会話で話題が転換していても、時刻の取り方に個人的なばらつきが見られた。これは、「うーん」や「ええと」のような、新しい発話が始まったのか決めかねるような表現などにより、話題の転換時刻にずれが生じているものと考えられる。この話題転換時刻のずれを補正するため、アフターコード方式を採用し、著者らは実験で用いた2種類の再生映像を再度見直し、同じ発話によって話題の転換が為されていると思われる話題転換時刻を1つのカテゴリに集約した。こうした手続きにより各再生映像とも、被験者によって挙げられた談話転換時刻が約80個のカテゴリにま

とめられた。以降ではこのカテゴリを、転換時刻カテゴリと呼ぶ。

被験者には各映像で話題が大きく転換した時刻を5個指定してもらった。この作業によって8人の被験者から5個ずつ抽出された40個の時刻を、話題境界候補点と呼ぶ。

各映像の、会議時刻における話題転換の大きい時刻として選択した被験者数の度数分布を図7に示す。縦軸は、その時刻が大きな話題転換点であると判断した被験者の人数を表す。

ここで、図7で縦軸の度数が2以上の転換時刻カテゴリに注目する。これは被験者8名中、2名以上がこの時刻で話題が大きく転換していると判断したことになる。各映像とも、この度数が2以上の転換時刻カテゴリは、10個あった。本論文では、この10個の転換時刻カテゴリを、被験者が選択した話題境界候補という意味で、境界選択時刻と呼ぶ。各々の候補の時刻は、カテゴリの開始時刻を採用する。

また、図中の点線は、本研究で提案する話題境界検出アルゴリズムによって算出された10個の候補点の時刻を表している。この10個の候補を、アルゴリズムで算出された話題境界候補という意味で、境界算出時刻と呼ぶ。

5.2 メモ選択に関する調査実験の概要と結果

会議中に記録したメモから再生場面を選択するとき、どのような場面から再生する傾向があるのかを調査する実験を行い、5.1の実験結果と比較した。

実験の設定は、実験5.1とほぼ同様であり、インタビューについてのみ違っている。被験者にはレポート作成終了後、会議記録時の作業者が書いた10個程度のメモを見てもらい、そのメモのうち、会議を通して重要だと思われるメモを5個選択してもらった。その5個について、1つずつ「記録者がこのメモを書いた意図を理解したいとき、あなたはどこから映像を再生しますか」という質問で、レポートで挙げてもらった話題転換時刻のうち1つの時刻を選択してもらった。

筑波大学工学システム学類の学生を中心に8名(男

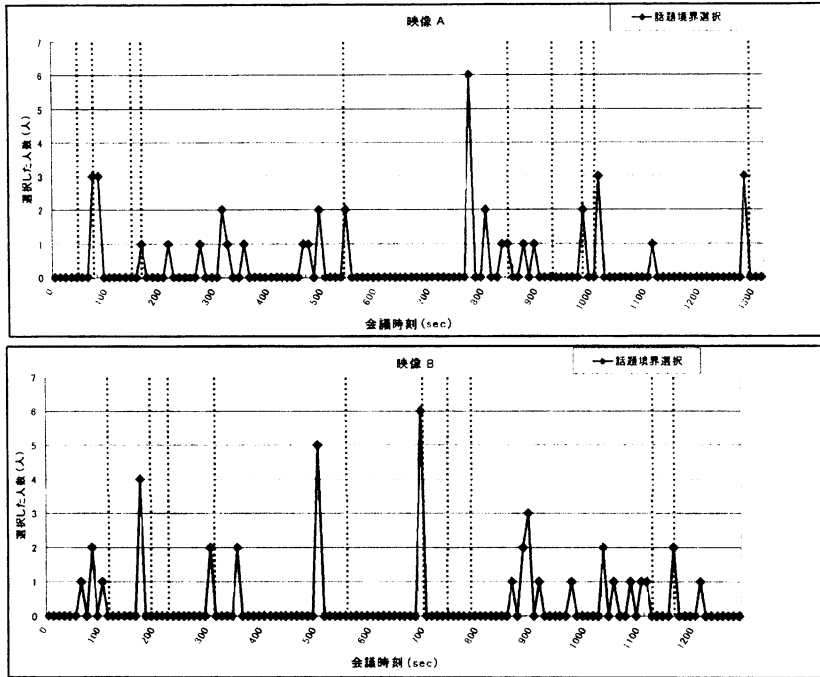


図7 話題境界候補として選択した時刻と人数

女各4名)の被験者に対して実験を行なった。再生映像は実験5.1と同じ2種類の映像を、4名(男女各2名)ずつ映像の順序を変えて行なった。上記の設定からこの実験では、レポートで記録した話題転換時刻から会議において重要なメモの再生候補点が選択される。8人の被験者から5個ずつ抽出された40個の時刻を、メモ再生候補点と呼ぶ。

図8は、実験5.1での図7にメモ再生候補として選択した時刻と人数の関係を点線で重ねたものである。なお、図7とは違い、図8では縦の点線は、記録者がメモを書き始めた時刻を指す。

5.3 実験の評価

実験5.1において、境界選択時刻と境界算出時刻を比較すると、各映像とも10個中3個は、結果が一致していると考えられる。その一方、50秒前後の差が存在する時刻が確認された。これは明らかに話題の転換が大きいとして算出された時刻が、被験者によって話題の転換が大きいと選択された時刻と一致していない。なぜこのような候補点が取られたか映像を観察してみたところ、議論が盛り上がり相手の発話に割り込むために声を大きくする場面や、メモを取ることに集中したために発話ポーズが長くなる場面が候補として挙がったことが確認できた。一方、実験5.2において、図8から、話題の境界を選択した人数が多い時刻では、メモ再生の時刻も多くの人数が選択している傾向があ

ることが伺える。

5.4 考察

本システムは、話題の境界を検出するアルゴリズムを既存の遠隔会議システムAgoraに組み込むことで、メモに関連性の高い場面の映像を提示することを目標とした。実験5.1では、本研究で提案する話題境界検出アルゴリズムが人間が選ぶ話題境界と一致するかを確認し、実験5.2では、話題の転換が大きい時刻とメモから再生したい時刻との関係を示した。

実験5.1の結果より、今回用意した2つの20分の会議映像では、10個の候補中3個が一致した。これは一見低い数値のように感じる。ただし、一致しなかった候補についても、小さな話題転換と判断できる箇所や議論の盛り上がりで声が大きくなった箇所など、会話の中の特徴点として有意な場所が含まれていた。そのため、長い会議映像から参照したい場面を検索する際、候補点の大半についてどんな内容の話し合いをしているのか判断するのに十分有効な候補であると予測できる。これについては今後、提案するシステムを利用した評価実験を行い、分析を通じて確認する予定である。

一方、実験5.2の結果より、話題の転換が大きい時刻とメモから再生したい時刻は比較的強い相関があることが示された。これにより本研究の目的である話題の境界の提示は、メモから映像を再生する方法として、

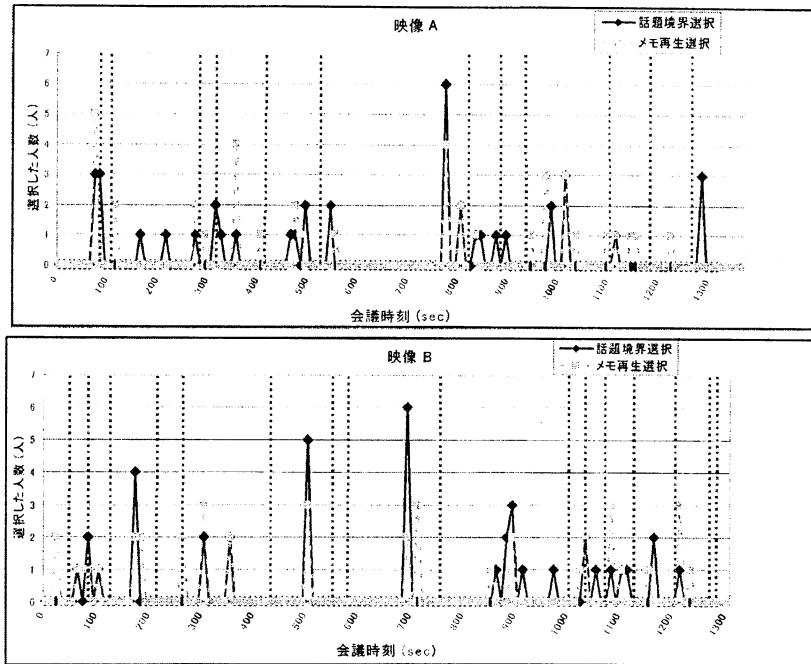


図 8 メモ再生候補と話題境界候補

有効であると考えられる。

今後は、同時発話などの他の言語的な要因を考慮した上で、話題境界を検出する精度を向上させるとともに、ユーザが参照したい場面を正確に提示できるシステムの開発をする。

6. おわりに

本稿では、話題の境界の提示という方法により遠隔会議の再生を支援しようと試みた。その結果、話題の境界の提示が有用であり、3割程度の正確さで話題の境界が提示される環境が実現した。

今後、ネットワーク基盤の充実と低価格化が進めば、より一層遠隔地間での会議が活発になることが期待できる。これにより、遠隔会議システムを利用する人々は専門的な知識を持った人だけでなく、機械操作に不慣れな人が使用することも考えられる。様々な人が容易に遠隔会議を利用し、会議の様子を記録・再生する環境では、氾濫した映像データの中から参照したい映像や場面が瞬時に提示されることが望まれるとともに、誰にでも使うことの出来る、「直感的で自然な」インターフェースが必要である。従って、今後は再生候補点の算出アルゴリズムの改良だけではなく、再生候補点の提示インターフェースも含めた改良を行っていく必要がある。

参考文献

- 1) 倉本到, 野田潤, 藤本典幸, 萩原兼一: 会合における防備録をもとに一次記録を検索参照する会合情報記録検索システム ReSPoM; 情報処理学会論文誌, Vol. 41, No. 10, pp. 2804-2813 (2000).
- 2) Sifelman, L., Arons, B., Schmandt, C.: The Audio Notebook; Proc. of CHI2001, pp.182-189 (2001).
- 3) 井上直人: 遠隔ミーティングの記録・再生を支援するシステムの研究; 筑波大学大学院博士課程システム情報工学研究科修士論文 (2004).
- 4) Koiso, H., Shimojima, A., and Katagiri, Y.: Collective signaling of informational structures by dynamic speech rate; Language and Speech, Vol. 41, Nos. 3-4, pp. 295-321 (1998).
- 5) 大久保崇, 菊池英明, 白井克彦: 音声対話における韻律を用いた話題境界検出; 情報処理学会研究報告音声言語情報処理, Vol. 2003, No. 124 (2003).
- 6) 山下淳, 葛岡英明, 山崎敬一, 山崎晶子, 加藤浩, 鈴木栄幸, 三木弘之: 相互モニタリングが可能な遠隔会議システム; 日本バーチャルリアリティ学会論文誌, Vol. 4, No. 3, pp. 495-504 (1999).
- 7) 井上直人, 葛岡英明, 鈴木栄幸: 遠隔ミーティングの記録・再生を支援するシステムの研究; ヒューマンインタフェースシンポジウム論文集 2003, pp. 745-748 (2003).