

会議途中参加支援を目的とした会議情報検索インタフェース - GijiLook の提案 -

犬 童 拓 也[†] 細 田 真 道[†]
中 山 彰[†] 岩 城 敏[†]

昨今、利用環境が整ったネットワーク型会議システムは、会議で相互にやりとりする映像、音声、テキスト、資料の情報がすべて再利用可能であることに利点がある。我々は、蓄積したマルチメディア会議情報を利用し、各メディアの持つ特徴量の関係から会議の構造を整理することで、会議途中参加者への効果的な情報提示や、会議参加者への振り返りを支援するシステムである”おっかけ会議システム”と称して提案している。本稿では、参加者の要求によって会議情報を検索するインタフェースである”GijiLook”を提案する。

GijiLook - A Search System from Conference Multimedia Data for Halfway-participant

TAKUYA INDO,[†] MASAMICHI HOSODA,[†] AKIRA NAKAYAMA[†]
and SATOSHI IWAKI[†]

To give a novel function to support halfway participation for conventional Internet video conference systems, we propose a flexible conference searching method “ GijiLook ” that can provide us answers for the participant 's individual questions and demands for the conference contents. The GijiLook is implemented on the RDBMS with time-series conference media feature measure tables that are defined for all users 'conference media such as voice, video, chat text and sharing documents. This paper describes the background of this work, GijiLook framework with some examples of conference media feature measure tables and the system implementation.

1. はじめに

インターネット接続環境の普及とブロードバンド化が進んだことにより、ネットワーク型会議システムを利用できる環境は整いつつある。

従来より、ネットワーク型会議システムを利用することで、移動時間の削減や会議スペースの縮小など、時間的、空間的なコスト削減が図れることが大きな利点として言われ、映像や音声品質の向上など実際の会議に近づける試みがなされてきたが普及には至っていない。

我々は、ネットワーク型会議システムの大きなメリットは、メンバの召集が簡単になり会議参加者のグルーピングの自由度が上ることにあると考えている。会議への参加が簡単になることで、時間が許す範囲で会議に参加したり、関心のある話題についての議論の場面だけ参加するといった自分にとって都合のよいマイペースな会議への参加方法が今後は望まれてくると予

想される。

このような新しい会議への参加方法に対して、我々は、蓄積されたマルチメディア会議情報を利用し、会議の要所や雰囲気を出出することで会議途中参加者への効果的な情報提示や振り返りを支援する”おっかけ会議システム”と称して研究を進めている¹⁾すなわち、“おっかけ会議システム”では、相互にやりとりする映像、音声、テキスト、資料等の情報がすべてデジタルデータとして蓄積されるネットワーク型会議システムの利点を活かし、それらを途中参加者の支援に最大限利用とすることを目指している。

なお、広義の「会議」は、人と人とのコミュニケーション全般を差すが、本研究においては最終的になんらかの合意や意思決定を目指し複数回の会合を行う一般的なタイプの会議を対象とする。

本稿では、参加者の要求によって会議情報を検索するインタフェースである”GijiLook”を提案し現状の実装について報告する。

[†] 日本電信電話株式会社 NTT サイバースペース研究所

2. 従来の会議情報検索・提示手法のサーベイと本研究のアプローチ

2.1 従来の手法

これまでも、会議情報の提示として発言構造を表示したり、途中参加支援を行うことを目的とした会議システムは存在した。

例えば、gIBIS²⁾では会議の情報を整理するために参加者の意図的なタグ付与による構造化を行った。同様に同期・非同期統合型会議システム ASSIST³⁾では、ツリー構造による会議の可視化も実装されており、部分的な会議映像の再生を可能としている。これらは、タグの精度という観点では信頼性が高いが、会議中に会議参加と並行して予め用意された枠組みに従った情報整理を参加者に強いることとなり、参加者の負担が大きいことが問題として指摘されている。

また、DYNAMITE⁴⁾⁵⁾では発言音声を中心とし、発言者が中心人物であるかどうかという発生状況を考慮したダイジェストを提示することでモバイルユーザの途中参加支援を目指した方式を提案しているが、ダイジェストの効果が会議の内容や利用者の好みに依存するため一概に評価することは難しいと述べられている。

2.2 分析

確かに、会議の情報を整理するためには、発言が最も重要な情報である。発言の意味情報に着目したアプローチを考える時、理想的には会議において参加者が発言する内容をプログラムが理解し、会議の構造を整理することが望まれる。しかし、プログラムが人間の自由な自然会話をリアルタイムに把握し、意味的な構造を把握することは現時点でも困難であり補助的な情報が必要となる。

発言内容以外の補助的な情報として、身振りやうなずきなど、非言語情報がコミュニケーションにおける重要な要素であることはこれまでも言われており、対面の対話における感情の変化が発話量やまばたきと相関があるとの報告もある⁶⁾。映像や音声の解析による非言語情報の抽出と会議構造の解析への利用は有用で手法であると考えられる。

更に、途中参加者が必要とする情報は、参加者毎に興味のあるポイントは異なることが想定される。会議には複数の話題があり、参加者それぞれに重要視している内容が違うからである。すなわち、一律に会議の内容を要約して提示することには理解の促進という観点から問題がある。

2.3 本提案のアプローチ

上記の分析より、会議の途中参加を目的とした情報

提示で短時間に会議の内容を把握してもらうためには、1) 無意識の情報である非言語情報を活用し会議の構造を把握し提示すること、2) 複数のメディアを活用し会議の雰囲気伝えること、3) 参加者それぞれに異なる興味のあるポイントを表示することが重要であると考えられる。

“おっかけ会議システム”においては、会議の内容を要約した文章を提示するのではなく、検索により自分の意思で記録された会議情報にアクセスすることを主な機能として提供する。検索によって、参加者それぞれに異なる会議中の気になるポイントを、会議のダイナミズムを感じながらできるだけ多くのメディアを用いて再生させる。検索においては、音声や会議資料のキーワードだけでなく、非言語情報である参加者の発言の音声的な特徴や、映像的な特徴、マウスポインタの動作、生体的な情報を利用し途中参加者の必要としている会議の一部と、全体のおおまかな構造を提示することを最終的な目標とする。

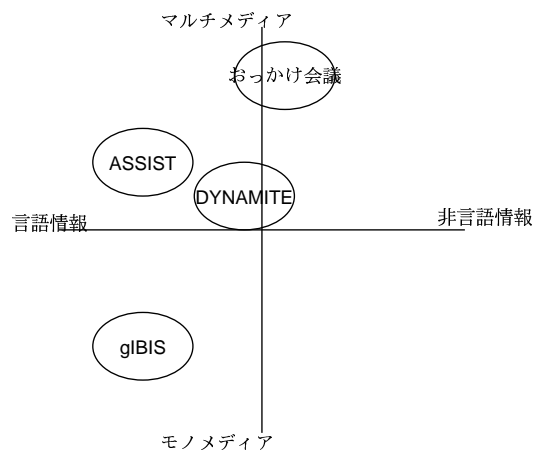


図1 手法の比較

これらの特徴について、利用できるメディアの数と会議情報整理に利用する情報の種別を軸に、過去の報告と本報告の手法について比較した結果を図1に示した。

3. GijiLook の提案

3.1 想定質問の整理

“GijiLook”を検討するにあたって、途中参加者がどのような情報を必要とするかを整理した。

途中参加者の想定質問としては、我々のグループで在宅勤務を想定して過去に実施した多地点の遠隔会議実験において収集したアンケート結果を利用する⁷⁾

以下に、アンケート結果から抽出した質問や要求の具体例をいくつか示す。

- 質問 1 途中参加者の会議中に最もあわれた単語は何か?
- 質問 2 チャットで一番使われた単語は何か?
- 質問 3 特定の人物が特定のキーワードを発したタイミングはどこか?
- 質問 4 特定の人物が「 」と言った後の参加者の様子を見たい
- 質問 5 特定の人物の語気が強まったシーンは?
- 質問 6 一番お喋りな人と寡黙な人はだれ?
- 質問 7 一番長く表示されていた資料はどれ?
- 質問 8 会議資料に「テレビ会議」と表示された時に「マルチキャスト」と発言した人はだれか?
- 質問 9 質問と回答のやりとりの場面が見たい
- 質問 10 一番盛り上がったところと、盛り下がったところは?

これらの疑問や要求は、このテキスト通りの自然言語の形でシステムへ入力し処理されるのが1つの理想であるが、現状の技術では当然難しい。そこでこれらをシステムが理解できる形式に変換あるいは翻訳する必要がある。

これらの想定質問には、単語と関連した質問が多く見られ、発話の音声認識結果やチャットおよび会議資料からの単語抽出を行い単語の一覧を作成する必要があると考えられる。想定質問への回答を検索する場合に、単語一覧以外に必要な情報を下表に整理した。

表 1 検索時に必要な情報

質問 1	単語別出現回数
質問 2	単語別出現回数, 出現メディア
質問 3	単語出現時刻, 発言者
質問 4	単語出現時刻, 発言者, 参加者映像
質問 5	単語出現時刻, 発言者, 音量
質問 6	参加者別総発言長
質問 7	資料表示開始/終了時刻
質問 8	資料表示開始/終了時刻, 資料中表示単語, 発言者, 発言時刻
質問 9	質問発生時刻, 回答発生時刻
質問 10	盛り上がり指数

このように、途中参加者の質問に答えるためには、単語の発生時刻だけでなく単語が発生した 5W1H 的な状況を考慮する必要がある。

質問 10 にある「盛り上がり指数」については、どのような特徴量を用いて定義するかは検証が必要であるため、「GijiLook」では様々な特徴量を検索に利用できる枠組みが必要となる。

3.2 データ構成

“GijiLook”において蓄積するデータとして、まず発言および資料、チャット中に出現する単語を抽出したインデックスが重要である。

その他、発言の開始時刻や終了時刻、音声や映像を解析したデータも蓄積する必要があるが、今後も増加、変更することが想定されるため、RDBMS によるテーブルとして整理しデータ構造の拡張性を確保する。

例えば、発言については図 2 のように、「発言音声テーブル」として、「発言発生時刻」「発言終了時刻」、「発言者 ID」、「平均ピッチ」、「平均音量」、「発言中の単語」で構成される

発生時刻	終了時刻	参加者ID	発話時間	ピッチ	音量	単語
0:34:56	0:37:22	2	2:26	110Hz	20dB	今日, 会議
0:44:56	0:47:22	1	2:26	120Hz	18dB	参加者, 数, 3人
0:54:56	0:57:22	3	2:26	130Hz	21dB	課題
...

図 2 発言音声情報テーブルの例

単語のインデックスおよび「発言音声テーブル」によって、質問 1, 2, 3, 4, 5 に必要なデータが格納される。

また、質問 6 に必要な参加者別の発言長も「発言音声テーブル」から発言者 ID 毎に発言終了時刻と発言発生時刻の差を加算することで算出可能である。

質問 7 に対しては、「共有資料操作テーブル」として「資料 ID」、「発生時刻」、「参加者 ID」、「操作種類」、「表示ページ」等を蓄積することで各資料の表示時間合計を算出し比較することで対応できる。質問 8 に対しては、「共有資料操作テーブル」より各資料の表示開始/終了時刻をとり、「発言音声テーブル」と比較して資料表示中に発言した参加者の一覧を作成することで対応できる。

発生時刻	参加者ID	操作種類	表示ページ	内容	単語	書き込み内容	書き込み単語
0:34:56	2	順めくり	7	課題として...	課題		
0:44:56	1	逆めくり	6	電子会議における...	電子会議		
0:54:56	3	書き込み	6	電子会議における...	電子会議	会議とは?	会議
...

図 3 共有資料操作テーブルの例

質問 9 に対しては、「発言音声テーブル」に含まれる発言の内、何が質問なのか回答なのかを判別する必要がある。この時、“GijiLook”では発言内容の解析ではなく、非言語情報を利用して判別を試みる。例えば、「発言音声テーブル」から発言中のピッチの変化を比較し、発言後半部分のピッチが平均より高い発言を質問口調の発言と判断し、質問発言の直後に発言者が変更しピッチの変動が小さい場合を回答とする等で検出が

可能であると考えている。

質問 10 に対しては、「盛り上がり指数」の定義が必要となり、今後のコーパス解析によってパラメータを考慮していく。例えば、会話における自然な手順を逸脱する場面を「盛り上がり」として定義することが考えられる⁸⁾。これは、相手の発言終了を待たずに割り込んで発言する場合や、全体を通して人の発話の声の状態が発言の密度が高い場合、参加者にとってなんらかの意思があり、会議において意味のある場面であるという仮定である。発言の重なりや密度は、「発言音声テーブル」を元に発言数、発言開始時間、発言終了時間、発言長を算出することで対応できる。更に、図 4、図 5 にある「顔画像情報テーブル」や「脈拍センサーテーブル」も利用し表情や脈拍における変化も利用可能である。

発生時刻	参加者ID	顔色	顔画像領域の中心座標	顔画像領域の大きさ	表情
0:34:56	1	R255,G200,B180	10,20	150	喜
0:34:56	2	R255,G200,B181	12,22	140	怒
0:34:56	3	R255,G200,B182	8,18	130	悲
0:34:57	1	R255,G200,B183	10,20	150	楽
0:34:57	2	R255,G200,B184	12,22	140	喜
0:34:57	3	R255,G200,B185	8,18	130	怒
0:34:58	1	R255,G200,B186	10,20	150	悲
0:34:58	2	R255,G200,B187	12,22	140	楽
0:34:58	3	R255,G200,B188	8,18	130	無表情
...

図 4 顔画像情報テーブルの例

発生時刻	参加者ID	脈拍	電気抵抗	α波レベル	β波レベル
0:34:56	1	80	800k	10	20
0:34:56	2	70	600k	11	21
0:34:56	3	90	1000k	9	22
0:34:57	1	80	800k	10	20
0:34:57	2	70	600k	11	21
0:34:57	3	90	1000k	9	22
0:34:58	1	80	800k	10	20
0:34:58	2	70	600k	11	21
0:34:58	3	90	1000k	9	22
...

図 5 脈拍センサーテーブルの例

3.3 検索実施例

検索クライアントの画面を図 6 に示した。会議の雰囲気や違和感なく伝えるため、画面構成は会議クライアントと検索クライアントを同様のものとした。

利用者は、まず検索クライアントで検索対象となるメディアを選択する。必要に応じて、対象ユーザの絞り込みも可能である。次に、検索キーワードを入力し検索する。

検索結果は、左側には検索結果とリンクされる会議中の特定の時点が赤線で表示され、利用者はその時点から会議を再生可能である。同様に複数の条件を利用する場合は「複合ボタン」を押し、条件入力部分を二

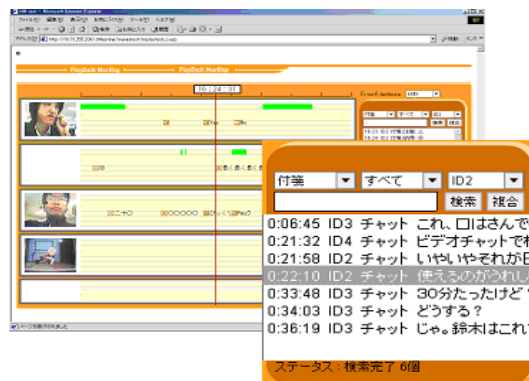


図 6 検索実施例

つ表示させる。そして、それぞれに必要な事項を入力し検索を行う。

例として、想定質問 8 における検索の実施例を示す。

まず、条件選択メニューで検索対象メディアとして「資料」を選択し、テキスト入力欄に「テレビ会議」と入力する。次に「複合」ボタンを押して条件欄を 2 個表示させ、複合選択の条件メニューで「かつ」を選択する。下の条件選択メニューで「発言」を選択し、テキスト入力欄に「マルチキャスト」と入力し検索ボタンを押す。

サーバ側では、資料中の単語一覧に「テレビ会議」を含む資料を抽出し、該当資料の表示開始/終了時間を保存する。次に、上記の資料表示時間内に発生した発言を抽出し、発言内の単語一覧を作成する。最後に、発言内の単語一覧から「マルチキャスト」を探し、発言者 ID を決定するとともに発言発生時刻を示す。

以上の処理により、想定質問 8 に対する回答を作成する。

現状の実装では、現在、発言を音声認識してテキスト化したデータと、会議中に利用した資料中のテキスト部分を用い、Chasen および Namazu によって文法解析とインデックス作成を行っているが、RDMS との連携が完了した場合、検索時にはユーザの入力した条件を SQL 文として作成して検索を行う。今後、選択できる条件が増えた場合は、図 7 のように複数の条件を選択する画面も想定される。

4. システム全体の構成と特徴

最後に「GijiLook」を含む会議システム全体の構成を図 8 示した。

本システムは、サーバ/クライアント型となっている。サーバは、映像音声通信を制御する会議サーバ、会議中の操作等を記録し検索用の下処理を行う検索

○誰が?	○いつ?	○なにで?	○なにを?	○内容は?
Aさん Bさん Cさん 全員 別条件	時刻指定 すべて 別条件 開始時刻 0:00:00 終了時刻 0:00:00	音声 画像 チャット 付箋紙 マウス キーボード 共有資料 ホワイトボード センサ すべて 別条件	単語 発言時間 ピッチ 音量 すべて 別条件	自由入力 最多 すべて 別条件 自由入力 文章入力

図 7 検索条件選択画面例

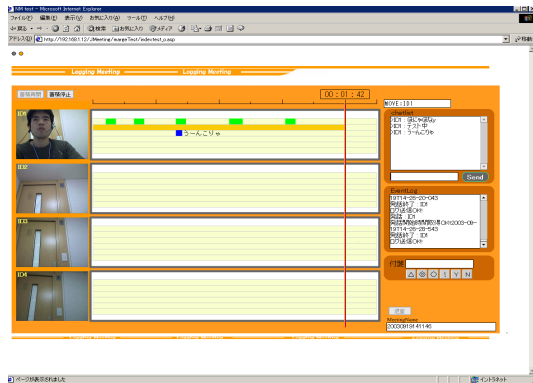


図 9 会議システム画面

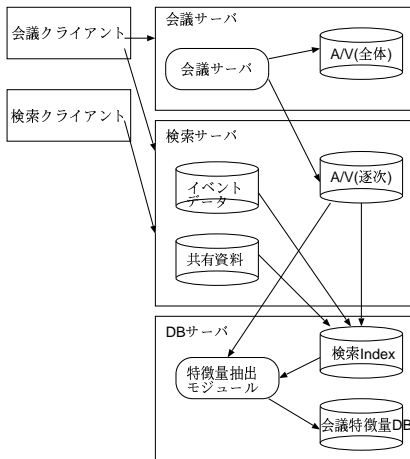


図 8 全体構成図

サーバから構成される。また、各種データを保存するデータベースサーバを付加する予定である。クライアントは、会議クライアントと検索クライアントに別れている。会議制御の基本的な機能は、Macromedia社のFlash Communication Server MX 1.5をベースとして利用し、4地点の同時利用が可能である。会議に利用可能なメディアは、音声、対地映像、共有資料、テキストチャット、会議の任意の時点でコメントを付与するアノテーションとなる付箋紙がある。付箋紙では、あらかじめ設定した や×等の記号を入力するボタンも用意した。

会議クライアント(図9)では、参加者の映像の横に発言状態が時系列に表示され、チャットや付箋紙のテキストも示される。これら、チャットおよび付箋紙のデータはテキストとして蓄積される。

映像と音声のデータは、会議進行中にも利用するため、を処理用と蓄積用の二系統に分けて蓄積することを考えている。その際、処理用の映像データは、シス

テム固有のフォーマットから AVI と WAVE 形式に変換する。変換されたデータを用いて、発話長やピッチ、発話の重なり具合、発言者の顔領域の動き等、会議における特徴量を抽出する会議特徴量抽出モジュールによって必要なデータを取得を経由し、会議特徴量 DB に格納することが可能とする。

共有資料やテキストチャット、アノテーションに含まれるテキストデータは、検索サーバにおいてインデキシングを行い、キーワード検索が可能な状態とする。

更に、他の生体情報取得センサ等との併用を考慮して、センサデータの取得と解析、蓄積に対応する。

5. 今後の課題

現在、会議コーパスの作成とその解析を進めている。今後は、このコーパスを利用し、会議特徴量の抽出度合を評価するとともに、会議参加者の音声や映像情報から得られる非言語情報について整理を進め、構造の解析に利用可能なパラメータを増していく予定である。

システムの有効性を検証するため“GijiLook”を利用する場合の途中参加者への支援度合について評価を行う。

会議特徴量の抽出度合の評価については、会議コーパスの作成において被験者が整理された会議内容の構造データを取得しており、参加者の主張する会議の構造と信号処理による特徴点の相関を検証する。

システムの有効性を評価する場合、参加者へのアンケートによる主観的な評価のみならず、タスク達成への寄与の度合や発言数の遷移などを定量的に測定できる環境を構築し、評価の指針としたい。

次に、評価を通じて途中参加や振り返りの支援に必要な項目を更に整理していく必要がある。特に、会議の振り返りを支援するためには会議全体を俯瞰し、短時間に会議内容を把握する必要がある。現時点では、

会議全体において参加者別の発話状態を時系列に一覧表示させている(図10)が、さらに縮退させた情報提示方法を行う。



図10 発話一覧表示

途中参加者の想定質問として、「今どうなっているのか」「何が決まったのか」といった抽象的な疑問にも対応する必要がある。前者の想定質問に対しては、例えば、内容に関わらず数分前に戻って会議を再生したり、話題の区切を抽出し直近の話題開始時点までを示すことで「今どうなっているのか」を示す方法が考えられる。また、後者に対しては、会話中の補助ツールとしてテキストチャットの利用方法を説明したところ、重要事項の共有メモとして活用する傾向が見られることから、話題終了近辺に発生したテキストチャット内容を表示する方法が考えられる。

6. まとめ

ネットワークを利用した会議システムにおいて、途中参加を行う場合にスムーズに議論に加わることができるよう、開催中の会議および過去の会議の情報を提示するインタフェース“GijiLook”について、構想と現状のインプリメントを報告した。

今後は評価を行い、会議への途中参加や振り返りを支援するシステムの拡充を行う。

音声や会議資料に含まれるテキストを対象として意味的な解析を行うだけでなく、映像や音声、その他参加者の行動に含まれる非言語情報を元に会議の構造をおおまかに把握できること、そして、その構造情報を利用することで会議への途中参加支援を目的とした情報提示システムの実現を行っていく。

参考文献

1) 中山, 細田, 犬童, 岩城: 多地点ビデオ会議コーパスに基づく会議途中参加支援機能の研究. GN49

(2003)

- 2) Conklin, J. and Begeman, M.L.: *gIBIS: A Hypertext Tool for Exploratory Policy Discussion*, Proc. of CSCW '88 (1988)
- 3) 田中, 福宿, 西堀, 勅使河原: 同期非同期統合型マルチメディア会議システム ASSIST におけるマルチメディア議事録の開発と評価, DICOMO'99
- 4) Yoshihiro Yamada, Ken Ohta, Tadanori Mizuno: Extracting and viewing information method for mobile tele-conference system, IEEE (1998)
- 5) 川口, 加藤, 石原, 酒井, 水野: 同期型会議へのスムーズな途中参加支援のための一方式, 情処論文誌, Vol.42, No.12 (2001)
- 6) 綿貫啓子, 外川文雄: 対話における感情の変化の解析, 音声言語情報処理 7-13 (1995)
- 7) 小林稔, 岩城敏: 多地点接続インターネット会議システムを用いた在宅勤務実験, 信学技報 OIS2002-11 (2002)
- 8) 日高ら: “音声強調に着目したマルチメディアコンテンツ要約技術”, FIT2002 予稿集, K-36 (2002)