

発話内容の視覚化による会議支援の検討

Study on meeting support by visualization of conversation contents

福島 伸夫[†] 西尾 浩一[†] 曾我 和由[†] 中村 浩司[†] 木全 英明[†]
 Nobuo Fukushima Koichi Nishio Kazuyoshi Soga Koji Nakamura Hideaki Kimata

1. はじめに

ビデオ会議・Web 会議等の会議市場は拡大するとみられており、近年の音声認識技術の精度向上に伴い、会議議事録の機械的書き起しを行う製品も出てきつつある[1]。現在の一般的な会議システム[2]では、遠隔地の相手とビデオ会議が行える、会議資料の共有/ペーパーレス会議が行える、録音・録画による会議記録の収集が行える、等の機能は実現されているが、それらは会議自体の開催を可能/便利にしたり会議記録を残すための機能であり、会議における議事内容そのものをより深く意味のあるものにする機能を実現することには至っていない。今回、進行中会議の発言内容を振り返る機能を実現することで、会議の全体俯瞰と要所の再確認が行える様になり、結果、会議の議論をより深く意味のあるものにするための実現を目指した。

2. 意味のある会議議論を実現するために

会議議論をより意味のあるものにするためには、現在行われている会議発言内容を正確に把握することが必要だと思われる。この会議発言の正確な把握を実現するためには、1) 会議中の発言をテキスト化するなど可視化できるようにすること、2) 各発言の発言者を識別できるようにすること、3) 目的とする発言を発言記録より容易に参照できるようにすることが必要だと思われる。ここで、1) の発言のテキスト化に関しては、一般的な音声認識エンジンの利用にて実現できるため、本論文では 2) の発言者の識別 と 3) の目的とする発言の参照方法 に関して検討を行う。

2.1 発言者を識別する手法の検討

会議発言における発言者を識別する手法としては、表 1 に示す手法が考えられる。案 1 の発言者毎に用意したマイク音量により発言者を特定する手法は、シンプルで実現可能ではあるが、発言者毎にマイクを設置する手間が煩雑であり、利用されるケースは限定的となってしまう。案 3 の集音するマイク自体が音の発せられた方向を検知し発言者を検知する手法は、会議卓中央部にマイクを 1 台設置するだけのため手間はかからないが、現在本機能を持つ市販マイクは存在しない。案 2 の顔画像の変化により発言者を識別する手法は案 3 同様、会議卓中央部に 1 台のカメラを設置するだけのため手間がかからなく、また近年の安価な全方位カメラの登場により手軽に実施することが可能となっている。よって今回、案 2 の手法により発言者識別の実現を試みる事とした。

2.2 目的とする発言の振り返り機能の検討

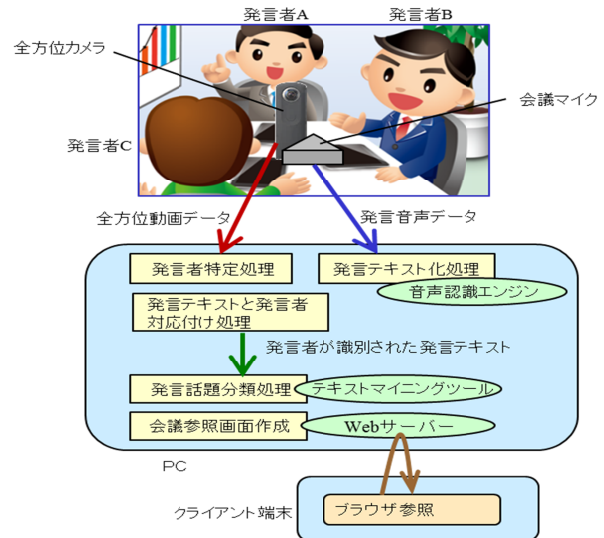


図1 試作システムの概要

表1 発言者特定手法の比較

手法	コスト	準備のための 手間	発言者 識別精度
【案1】 発言者毎にマイクを用意し、各マイクの音量変化により発言者を特定	△ マイクを発言者毎に用意する必要があり、コストが高い	× 発言者毎にマイクを設置する必要があり、手間がかかる	○ マイク音量は正確に測定できるため、発言者識別精度は高い
【案2】 顔画像(口部)の変化により発言者を特定する	○ 安価な全方位カメラの登場により、低コストでの実現が可能	○ 中央部に1台のカメラを設置するだけのため、手間は少ない	△ 俯いた場合等、正面画像の取得ができない場合は識別不能
【案3】 集音するマイクが音が発せられた方向を検知し、発言者を検知する	△ 特殊なマイクを導入する必要がある	○ 中央部に1台の専用マイクを設置するだけのため手間は少ない	× 音が反響するため、正確な発言者識別は難しい

現在多くの会議システムで用いられている録画・録音による会議記録収集では、参照したい内容/発言者が限定的であったとしても、話題内容や発言者による頭出しやフィルタリングが行えないため、録画時間通しての参照が必要であった。特に、会議中に経過議事の参照を行う事に関しては、会議に参加しつつ時間のかかる再生・視聴による参照が必要となるため、通常実施されることはなかった。

ここで、会議発言がテキスト化され、また 2.1 章に記した発言者識別情報が付与される状況になれば、各発言者毎に発言をテキストにて参照出来るようになる。更にこの発言テキストにクラスター分析(テキストマイニング処理)を施すと、目的とする話題分類に限定した参照が出来るようになる。つまり、発言者と話題分類という2つの強力なフィルタリング機能を利用して可視化された発言テキストの参照ができるため、会議中における経過議事の確認も容易に行えるようになる。

[†] NTT コムウェア (株) 品質生産性技術本部研究開発部
 Research and Development Department, Core Technology,
 Quality Management and Engineering Division,
 NTT COMWARE CORPORATION

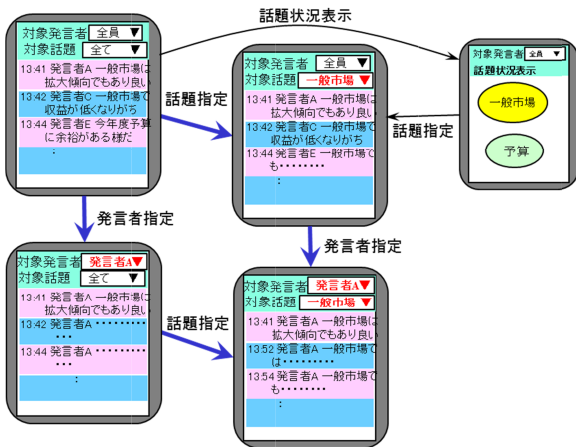


図 2 クライアント端末の議事参照画面

3. 試作システムの概要

試作したシステムの概要を図 1 に示す。会議卓中央部に全方位カメラと会議マイクを設置し、全方位カメラは参加者全員の顔が写っている全方位動画データを収集し、会議マイクは参加者全員の発言音声データを収集する。発言音声データは音声認識エンジンを用いてテキスト化すると同時に発言時刻の記録を行い、全方位動画データは画像分析を行い各時刻における発言者情報(発言者が誰であるか)を算出する。各発言テキストの時刻に応じて発言者情報を付与することで、発言者が識別された発言テキストが作成される。この発言テキストは更にテキストマイニングツール [3]を用いて話題分類処理が行われる。クライアント端末からは、議事参照を行う Web アクセスが行われる。

クライアント端末における議事参照画面を図 2 に示す。発言テキストが発言者および話題分類によりフィルタリングされて参照出来る様になっている。

4. 実用化に向けての課題

4.1 会議音声認識に適したマイク

今回、会議卓中央部に設置したマイクにて音声集音を行うが、この形態で音声認識(テキスト化)の精度が低くなる事象が発生した。原因追求のため発話者とマイクの距離を 10 人程度の会議を想定して 1.5m とした場合の各種マイクの音声認識精度の確認を行った(図 3)。

結果、指向性の低いマイクでは音声認識精度が低く実用に耐えられない状況であったが、指向性の有るマイクでは音声認識精度の低下が少なく、実用は可能だと思われた。ここで、従来からある指向性マイクでは、発言を行う発言者の方向に都度マイクを向け直す必要があるため、装置固定のままソフト的に発言者の方向に指向性を向けるマイクが音声認識を行う会議マイクとして有用と思われる。

4.2 画像認識による発話者識別

全方位動画画像による発話者識別は、機能的に実現することは可能であったが、実用的なレベルの精度を確保するには至らなかった。課題として、魚眼動画周辺部における整形処理(Calibration)や発話者識別判断を行う画像処理法に関して更なる検討を行うことが必要である。

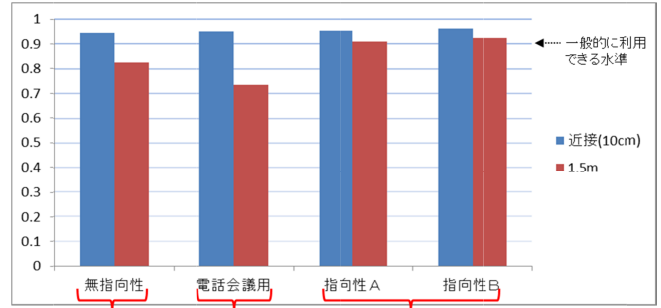


図 3 各種マイクにおける音声認識精度

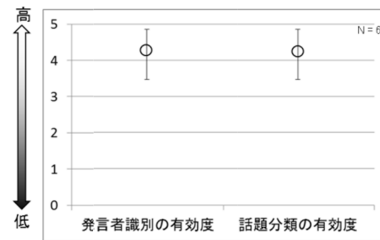


図 4 試作システムに対する期待度

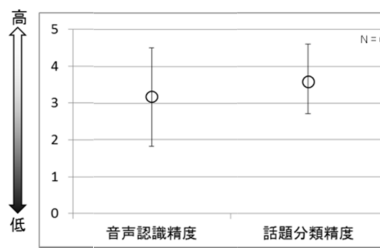


図 5 試作システムの精度

5. 本システムの評価

複数の被験者に試作システムを利用してもらい、アンケートに基づく主観評価を行った。

発言者識別と話題分類のフィルタリング機能を持つ発言記録の参照に関しては、有効度が高いとの回答が多く、高い期待度が得られた(図 4)。

試作システムの精度に関しては、指向性を持つマイクを用いるなど音声認識精度の向上を図ったが、被験者による認識精度のばらつきが多く、認識精度が低くなってしまったケースも見受けられた。また、音声認識精度が低くなった場合に、話題分類精度も引きずられて低下する状況が見受けられた(図 5)。

6. おわりに

発言内容の視覚化による会議支援の検討を行った。実用化に向けては、全方位動画の画像処理、会議音声の認識精度を更に向上させる検討が必要である。

参考文献

[1] 会議録作成システム VoiceAir, NTT 東日本, <http://www.ntt-east.co.jp/business/solution/minutes/>
 [2] MeetingPlaza, NTT-IT, <http://www.meetingplaza.com/>
 [3] 樋口耕一, 社会調査のための計量テキスト分析, ナカニシヤ出版, 2014