

平成21年6月22日現在

研究種目：基盤研究(A)

研究期間：2006～2008

課題番号：18200007

研究課題名（和文） マルチメディア型会議録の構造化に関する研究

研究課題名（英文） A study on the structuring of multi-media meeting data

研究代表者

浅野 太 (ASANO FUTOSHI)

独立行政法人産業技術総合研究所・知能システム研究部門・グループリーダー

研究者番号：00231895

研究成果の概要：本研究では、複数のマイクロホンと全方位カメラとからなる入力装置により会議内容を収録したマルチメディアデータ（映像・音声）から、いつ、だれが、どんな発言をしたかという情報を、音源定位・音源分離や音声認識技術などを用いて自動推定して、会議の構造を視覚化するシステムを開発した。このシステムにより、キーワードを含む場面を簡単に検索・再生し、会議の概要を短時間で把握できるようになる。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2006年度	10,200,000	3,060,000	13,260,000
2007年度	13,000,000	3,900,000	16,900,000
2008年度	7,900,000	2,370,000	10,270,000
年度			
年度			
総計	31,100,000	9,330,000	40,430,000

研究分野：総合領域

科研費の分科・細目：情報学・メディア情報学・データベース

キーワード：マルチメディア、会議録、ウェブ・コンテンツ、マイクアレイ、全方位カメラ、音源定位、音源分離

## 1. 研究開始当初の背景

様々な種類の会議が、日々多数行われている。このうち、議会などの大規模な会議は、人手により文章としての議事録（テキスト型コンテンツ）が作成されているが、中小規模の会議では、人手による議事録作成はコストに見合わない。このような会議では、会議の内容を後に残す手段として、ビデオ

カメラなどにより、音声と映像を記録する手段（オーディオ・ビジュアル型コンテンツ）が取られることが多いが、会議内容を把握するためには収録内容を最初から最後まで聞く必要があり、きわめて効率が悪い。そこで、より短時間で所望の情報にアクセスできるようなマルチメディア会議録の作成が強く求められている。また、中小規模の会議は、複数の人間の間のインタラクシ

ョン研究のターゲットとしても重要であり、近年注目を集めつつある

## 2. 研究の目的

本研究では、会議の内容を音声と映像で収録し、収録内容を解析・構造化して、迅速に検索・閲覧できる会議録コンテンツを自動的に作成するシステムの開発を目的とする。この目的の達成のためには、次のような技術課題が挙げられる。(1) 発言者一人一人にマイクロホンを装着するなどの煩わしい作業のない会議収録デバイスの開発、(2) 収録内容から、だれが、いつ、どのような発言をしたかの情報を自動推定する方法の開発、(3) 会議の概要をわかりやすく視覚化することのできる表示・閲覧方法の開発。

## 3. 研究の方法

(1)収録デバイスの作成：カメラアレイとマイクアレイを用い、会議参加者にマイクを装着することなく収録を行うことのできるデバイスを作成する。(2)会議録データベースの構築：制作した収録デバイスを用いて、小規模の会議録を収録し、ラベルを付与して、実験を行うためのデータベースを作成する。(3)発話イベントの検出・分離：音源定位・分離技術を用いて、発話イベントの検出・分離を行うアルゴリズムを開発する。また、検出した発話イベントに対して自動音声認識を施し、キーワードを抽出するソフトウェアを開発する。(4)ブラウザ及び統合システムの開発：検出した発話イベントの情報に基づき、会議録を効果的に閲覧するブラウザ及びこれを収録・解析システムと統合したシステムを開発する。(5)評価：システムの評価実験を行う。

## 4. 研究成果

(1)収録デバイスの作成：カメラアレイとマイクアレイを用いた収録デバイス MArc-II、MArc-III を試作した(図1参照)。このデバイスにより、会議参加者ごとにマイクロホンを装着することなく、2m程度まで離れた話者の音声・映像を収録することができる。

(2) 会議録データベースの構築：マーケットリサーチで用いられるグループインタビュー（1回90分程度）を4回収録した。この収録内容をデータベースとして利用するため、人手による書き起こしを施し、発話開始/終了時間を付与した CSJ フォー

マットにより記録した。



図 1:収録用デバイス MArc-III

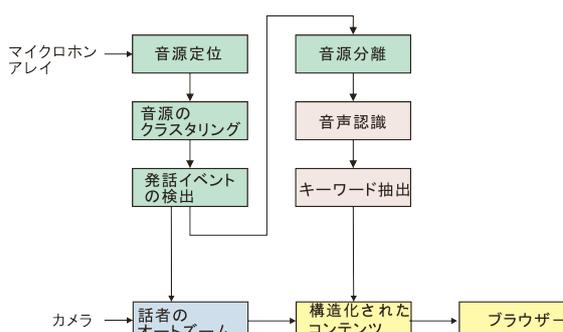


図 2 会議録コンテンツ作成の概要

### (3)発話イベントの検出・分離：

主成分分析・サブスペース法による音源定位・適応ビームフォーミングを用いて発話イベントを検出する手法を開発した。この手法により、音声が重畳した発話イベントでも、個人毎の発話に分離することができる。この成果は、国際学会 Interspeech2006 において発表し、EURASIP Journal on Audio, Speech, and Music Processing 誌に原著論文として掲載された。発話イベント検出結果に基づき、マルチメディア会議録コンテンツを作成するソフトウェアを開発した。ソフトウェアの概要を図2に示す。

### (4)ブラウザ及び統合システムの開発：

発話イベントを記述する言語 MADL、及びこれに基づいてマルチメディア会議録を効率的に閲覧することのできる MArcBrowser を開発した。図3に、開発した MArcBrowser を示す。左上のパネルでは、発話者が自動的にズームされる。左下のパネルには、発話イベントが表示され、これをクリックすることにより、発話イベントの再生が可能である。右中央のパネルには、頻度の高いキーワードが表示さ

れ、キーワードを含む発話イベントが右下のパネルに表示される。これらを用いて、効率的に所望の場面の検索・閲覧をすることができる。また、収録・解析を行うソフトウェアと統合し、MArcシステムを構築した。



図 3 会議録コンテンツを閲覧する  
MArcBrowser

#### (5) 評価

開発したシステムを、企業内の開発会議に適用し、キーワード認識率などの客観評価、及びアンケート調査（主観評価）を実施した。この結果、音声認識システムの言語モデルを適応しない場合は 41.2%、言語モデルを適応した場合は 57.2%のキーワード認識率が得られた。また、アンケート調査の結果から、映像・音声の検索は概ね良好であったが、キーワードの選定、発話者数の推定などに問題点もあり、今後の改善が望まれることもわかった。

#### 5. 主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 12 件）

① Angela Quinlan and Futoshi Asano, “Tracking a varying number of speakers using particle filtering,” Proc. ICASSP 2008, pp.297-300, 2008, 査読有

② Y. Matsusaka, et al., “Unsupervised clustering in Multimodal Multiparty analysis,” Proceedings on the International Conference on Language Resources and Evaluation, Workshop on Multimodal Corpora, pp.27-31, 2008, 査読有

③ F. Asano, K. Yamamoto, J. Ogata, M. Yamada, M. Nakamura, “Detection and

Separation of Speech Events in Meeting Recordings Using a Microphone Array,” EURASIP Journal on Audio, Speech, and Music Processing, Volume 2007, Article ID 27616, 2007, 査読有

④ F. Asano, J. Ogata, “Detection and Separation of speech events in meeting recordings,” Proc. Interspeech 2006, pp. 2586-2589, 2006, 査読有

⑤ J. Ogata, F. Asano, “Stream-based classification and segmentation of speech events in meeting recordings,” Proc. MRCS2006, pp.793-800, 2006, 査読有

〔学会発表〕（計 4 件）

① 松坂要佐, 浅野太, “facekit.net: Webから使えるカスタマイズ可能な顔認識器”, 情報処理学会 インタラクシオン 2007, pp. 83-84, 2007、査読無

② 浅野太, マイクロホンアレイとカメラアレイを用いた会議録コンテンツ作成システムについて, 音声ドキュメント処理ワークショップ, pp.147-152, 2006, 査読無

〔図書〕（計 0 件）

〔産業財産権〕

○出願状況（計 1 件）

産業財産権の名称：全周映像のための表示技術

発明者：松坂要佐、浅野太

権利者：産業技術総合研究所

産業財産権の種類、番号：特願2008-194305

出願年月日：H20/07/29

国内・外国の別：国内

○取得状況（計 0 件）

〔その他〕

① 浅野太, 松坂要佐, 関口智嗣, 大嶋裕子 “会議の映像・音声データから自動的に会議録コンテンツを作成するシステム”, 記者会見（プレスレク）、2008/10/14

## 6. 研究組織

### (1) 研究代表者

浅野 太 (ASANO FUTOSHI)

産業技術総合研究所・情報技術研究部門・  
研究グループ長・00231895

### (2) 研究分担者

麻生 英樹 (ASOH HIDEKI)

産業技術総合研究所・情報技術研究部門・  
主任研究員

研究者番号：10344194

河本 満 (KAWAMOTO MITSURU)

産業技術総合研究所・情報技術研究部門・  
研究員

研究者番号：10300865

緒方 淳 (OGATA JUN)

産業技術総合研究所・情報技術研究部門・  
研究員

研究者番号：10392599

松坂 要佐 (MATSUSAKA YOSUKE)

産業技術総合研究所・情報技術研究部門・  
研究員

研究者番号：10343625

### (3) 連携研究者