

「共生社会に向けた人間調和型情報技術の構築」

平成 21 年度採択研究代表者

平成 21 年度
実績報告

河原 達也

京都大学 学術情報メディアセンター ・ 教授

マルチモーダルな場の認識に基づくセミナー・会議の多層的支援環境

§ 1. 研究実施の概要

本研究では、人間の知的活動の源泉ともいえる音声コミュニケーションをマルチモーダルな観点で分析・モデル化した上で、セミナー・ポスター発表及び会議を対象として、リアルタイムに支援したり、効果的なアーカイブ化を行うための情報環境を構築する。主な話者の発話内容を音声認識して言語解析を試みるという従来のアプローチ(コンテンツに基づく処理)だけでなく、視線やあいづち・うなずきなどの聴衆の反応に着目した新たなアプローチ(インタラクションに基づく処理)を導入する。

初年度の平成 21 年度では、研究環境の整備につとめるとともに、マルチモーダルなコーパスの収集及びアノテーションを開始した。上記の研究目標を実現するための要素技術に関して、研究者間で意見交換・議論を行いながら、問題点を整理し、各要素技術の研究に着手した。具体的には、ポスター発表に基づく会話(ポスター会話)を多数のセンサを用いて収録し、マルチモーダルな情報のアノテーションを行った。また、セミナー室・ポスターボード向けのカメラやマイクロフォンアレイの設計を行った。さらに、音声コミュニケーションにおける聴衆の反応の認識やジェスチャーの認識に関する基礎的な研究を行った。

§ 2. 研究実施体制

(1) 京大グループ

① 研究分担グループ長:河原 達也 (京都大学、教授)

② 研究項目

マルチモーダルな場の認識に基づくセミナー・会議の多層的支援環境

具体的には、音声認識、映像処理、反応の認識、質問応答・情報推薦

(2) 奈良先端大グループ

① 研究分担グループ長: 鹿野 清宏 (奈良先端科学技術大学院大学、教授)

② 研究項目

セミナー・会議のための音響・音声処理

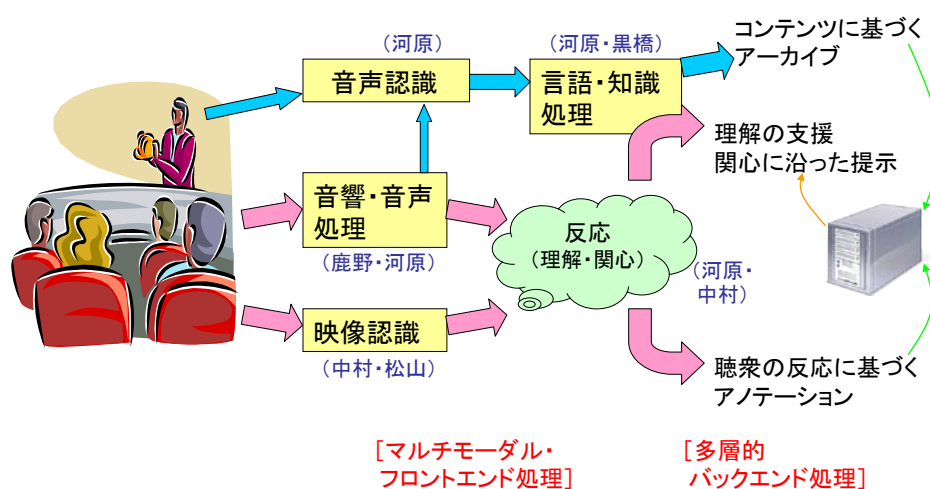
§ 3. 研究実施内容

(文中に番号がある場合は(4-1)に対応する)

本研究では、人間の知的活動の源泉ともいえる音声コミュニケーションをマルチモーダルな観点で分析・モデル化した上で、セミナー・ポスター発表及び会議を対象として、リアルタイムに支援したり、効果的なアーカイブ化を行うための情報環境を構築する。主な話者の発話内容を音声認識して言語解析を試みるという従来のアプローチ(コンテンツに基づく処理)だけでなく、視線やあいづち・うなずきなどの聴衆の反応に着目した新たなアプローチ(インタラクションに基づく処理)を導入する。知能化したセミナー室やポスターボードを構築し、定例のセミナーやポスター発表会で実証実験を行う。さらに、音声認識システムについては、衆議院の次期会議録作成システムで運用して得られる大規模なデータ・知見をフィードバックすることで、音響モデル・言語モデル(辞書)の高精度化を行い、幅広い話し言葉音声の書き起こしに供することができるようにする。

本研究の概要を図1に示す。

図1: 本研究の処理の概要



初年度の平成 21 年度では、研究環境の整備につとめるとともに、マルチモーダルなコーパスの収集を開始し、アノテーションを設計した。上記の研究目標を実現するための要素技術に関して、研究者間で意見交換・議論を行いながら、問題点を整理し、各要素技術の研究に着手した。具体的には以下の通りである。

○ 実験環境の整備とコーパス収集分析

実験環境の整備を行うとともに、コーパスの収集に着手した。

- (1) 実験室環境の整備: セミナー・ミーティング、電子掲示板によるポスター発表を行える環境を整備した。マイクやカメラなどのセンサ群を設計・購入した。
- (2) コーパスの収集とアノテーション: 多数のセンサを配置した IMADE ルームにおいて、新たに8セッションのポスター会話の収録を行った(図 2 参照)。その上で、発話の情報に加えて、あいづちやうなずきなどの非言語情報も含めたアノテーションを行った。
- (3) マルチモーダル情報の分析・抽出と認知状態との関連付け: 非言語情報を抽出するための物理的特徴量と認知状態との対応付けに関して検討を行った。

○ マルチモーダル認識及び情報支援に関する研究開発

以下の要素技術の研究開発に着手した。図1に研究全体の処理における位置づけと分担を示している。

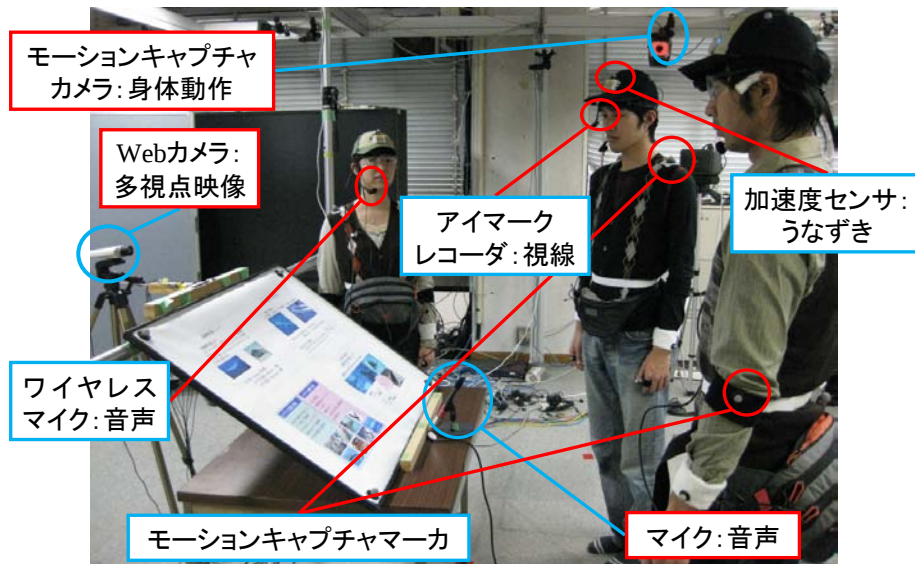
- (4) 音声認識: 国会討論の音声認識モジュール(音響モデル・言語モデル・単語辞書)を衆議院の会議録作成システムに向けて完成させるとともに、長期的な運用にむけてモデル更新を容易にする方法を研究した。
- (5) 書き起こしの整形及び構造抽出: 話し言葉にはフィラーや言い淀みが多いので、整形を行う方法を研究した。統計的機械翻訳のモデルに基づく手法を実装し、国会での発言と会議録を照合することで評価した。
- (6) 聴衆を対象とした音響・音声処理: 遠隔マイクを通して収録した音声に含まれる残響やミュージカルノイズを抑圧する処理の研究を行った。また、入力音声から発話区間を頑健に推定する方法の研究を行った。さらに、会話音声からあいづちや笑い声を頑健に検出する方法を研究した。
- (7) 聴衆を対象とした映像処理: 複数のカメラ映像を用いたジェスチャー認識・理解の基礎的研究として、3次元表面形状間の密な対応付けを計算する新たな手法を考案し^[文献 3]、3次元形状間の類似性尺度に関する検討を行った^[文献 4]。
- (8) 聴衆の反応の認識: 会話音声におけるあいづちと聴衆の興味の関連について調査を行った。その結果、特定の音韻・韻律的パターンを持つあいづちが興味の度合いと関連があることを特定した^[文献 1]。また、会話や会議のコミュニケーション状態をスナップショットとして記録する手法について検討を行った。

- (9) 質問応答及び情報推薦:話題や文脈に応じて関連した情報を検索し、わかりやすく提示する方法を研究した。音声認識誤りや検索の曖昧性を考慮した最適な対話戦略を考案した。また、専門用語などの未知語を頑健に獲得する方法を研究した^[文献2]。

○ セミナー・会議の支援システムの構築と運用

- (10) 衆議院での音声認識システム:衆議院の新会議録作成システムにおいて、音声認識の試験評価を行った。その結果、文字正解精度で85%を実現した。

図2 ポスター会話収録の様子



§ 4. 成果発表等

原著論文発表

●論文詳細情報

- [1] T.Kawahara, Z.Q.Chang, and K.Takanashi.
Analysis on prosodic features of Japanese reactive tokens in poster conversations.
In Proc. Int'l Conf. Speech Prosody, (accepted), 2010.
- [2] Y.Murawaki and S.Kurohashi.
Online Japanese unknown morpheme detection using orthographic variation.
In Proc. Int'l Conf. Language Resources and Evaluation (LREC), (accepted), 2010.

- [3] T.Tung and T.Matsuyama.
Dynamic surface matching by geodesic mapping for 3D animation transfer.
In Proc. IEEE-CVPR, (accepted), 2010.
- [4] P.Huang, T.Tung, S.Nobuhara, AHilton, and T.Matsuyama.
Comparison of skeleton and non-skeleton shape descriptors for 3D video.
In Proc. Int'l Sympo. 3D Data Processing, Visualization and Transmission (3DPVT'10),
(accepted), 2010.
- [5] T.Kawahara.
New perspectives on spoken language understanding: Does machine need to fully
understand speech?
In Proc. IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)
(invited paper), pp.46--50, 2009.
doi:10.1109/ASRU.2009.5373502