「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

H24 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. 研究実施体制

(1)京大グループ
①研究代表者: 河原 達也 (京都大学 学術情報メディアセンター・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

(2)奈良先端大グループ
①主たる共同研究者: 鹿野 清宏 (奈良先端科学技術大学院大学 情報科学研究科・教授)
②研究項目
セミナー・会議のための音響・音声処理

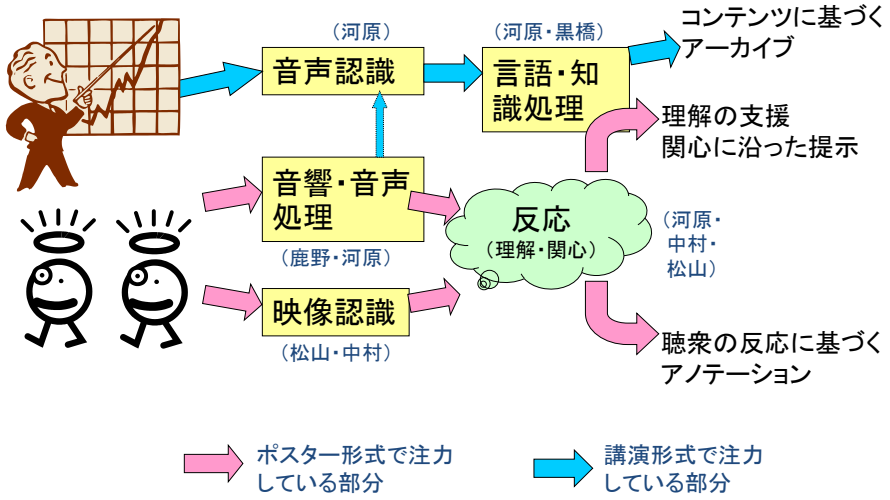
§2. 研究実施内容

(文中に番号がある場合は(3－1)に対応する)

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、効果的なアーカイブ化を行ったり、リアルタイムに支援するための情報環境を構築する。まず、学術的なイベント等で一般的になっているポスター形式の発表を対象に、多様なセンサを備えた大型液晶ディスプレイによる「スマートポスターボード」の設計・実装を行っている。本研究では、音声・音響・映像に関する研究者が結集して、視線やあいづちなどの聴衆の反応に着目したアプローチを提案・実装し、学会等でデモが行えるようにする。次に、講演形式のセミナーを対象に、音声認識による字幕付与と、言語解析・知識処理による用語解説を目指して研究を行っている。

本研究の概要を図1に示す。

図1　本研究の処理の概要



平成24年度は、これまでに研究開発してきた要素技術を統合して、スマートポスターボードシステムの構築を本格的に行い、研究会やオープンラボにおいて実演し、実際の会話データを収集した。また音声認識技術については、衆議院の会議録作成システムでの運用に加えて、京都大学OCW(OpenCourseWare)の講演に対する字幕付与への展開も行った。

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**

・コーパスの収集とアノテーション：スマートポスターボードを用いて、ポスターセッションの会話データを収録した。書き起こしや視線情報などのアノテーションを行った。

○　**マルチモーダル認識及び情報支援に関する研究開発**

音声認識：セミナー・講演などの音声認識のために、話者や話題に対してモデルを適応する方法を研究した[文献 10,11]。京都大学 OCW で配信されている iPS 細胞研究所の公開シンポジウム講演などを対象に評価を行った。

・書き起こしの整形及び構造抽出：音声認識結果に基づいて字幕を生成するための整形の方法を検討するとともに、そのための専用エディタの設計・開発を行った[文献 1,2]。

・聴衆を対象とした音響・音声処理：マイクロフォンアレイを用いて、遠隔話者の発話を分離・強調し、各話者の発話区間を検出する方法の研究を進めた[文献 5、22]。スマートポスターボードで収録した会話データを対象に評価を行った。

・聴衆を対象とした映像処理：カメラを用いて、聴衆の位置とふるまい（視線など）を検出する方法の研究を進めた[文献 31]。スマートポスターボードで収録した会話データを対象に評価を行った。

・聴衆の反応の認識：視線やあいづちなどの情報に基づいて、聴衆の興味・理解度を推定する方法を研究した。

・質問応答及び情報推薦：セミナー・講演で話される専門用語を抽出し、解説や関連情報を提示するシステムを実装した[文献 32]。

○　**セミナー・会議の支援システムの構築と運用**

・衆議院での音声認識システム運用：衆議院の会議録作成システムの運用において音声認識の評価とモデルの更新を行った。文字正解率で 90%を達成した[文献 15]。

・電子掲示板を用いたポスター発表：上記の要素技術を統合し、スマートポスターボードの本格的な構築を行い、下記の公開シンポジウムやオープンラボなどにおいて実演を行った。また、本システムについて国際会議の基調講演等で紹介を行った[文献 16]。

・セミナーの支援とアーカイブ：京都大学 OCW で配信されている iPS 細胞研究所の公開シンポジウム講演に字幕を付与した。また、本研究で主催した「聴覚障害者のための字幕付与技術」シンポジウムの講演において、音声認識を用いた字幕付与の実演を行った。

2012 年 4 月 1-2 日に京都大学において、本 CREST 領域に関する公開の国際シンポジウム CREST Symposium on Human-Harmonized Information Technology を主催した。自身の研究・プロジェクトに関する報告に加えて、世界的に著名な研究者による招待講演で構成し、参加者は2日間の合計で 132 名と大変盛況であった。

§3. 成果発表等

（3－1） 原著論文発表

● 論文詳細情報

[1] 秋田祐哉, 河原達也.
講演に対する読点の複数アノテーションに基づく自動挿入.
情報処理学会論文誌, Vol.54, No.2, pp.463--470, 2013.

[2] G.Neubig, Y.Akita, S.Mori, and T.Kawahara.
A monotonic statistical machine translation approach to speaking style transformation.
Computer Speech and Language, Vol.26, No.5, pp.349--370, 2012. (DOI: 10.1016/j.csl.2012.02.003)

[3] 三村正人, 河原達也.
会議音声認識における BIC に基づく高速な話者正規化と話者適応.
電子情報通信学会論文誌, Vol.J95-D, No.7, pp.1467--1475, 2012.

[4] 真嶋温佳, 藤田洋子, トーレス・ラファエル, 川波弘道, 原直, 松井知子, 猿渡洋, 鹿野清宏.
音声情報案内システムにおける Bag-of-Words を用いた無効入力の棄却.
情報処理学会論文誌, Vol.54, No.2, pp.443–451, 2013.

[5] R.Miyazaki, H.Saruwatari, T.Inoue, Y.Takahashi, K.Shikano, and K.Kondo.
Musical-noise-free speech enhancement based on optimized iterative spectral subtraction.
IEEE Trans. Audio, Speech & Language Processing, vol.20, no.7, pp.2080-2094, 2012. (DOI: 10.1109/TASL.2012.2196513)

[6] T.Tung and T.Matsuyama.
Topology Dictionary for 3D Video Understanding.
IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI), Vol.34, No.8, pp.1645-1657, 2012. (DOI: 10.1109/TPAMI.2011.258)

[7] Z.Yu, Z.Yu, X.Zhou, C.Becker, and Y.Nakamura.
Tree-based Mining for Discovering Patterns of Human Interaction in Meetings.
IEEE Trans. Knowledge and Data Engineering, Vol. 24, No.4, pp. 759--768, 2012. (DOI: 10.1109/TKDE.2010.224)

[8] K.Yoshino, S.Mori, and T.Kawahara.
Language modeling for spoken dialogue system based on filtering using predicate-argument structures.
In Proc. COLING, pp.2993--3002, 2012.

[9] C.Lee and T.Kawahara.
Hybrid vector space model for flexible voice search.
In Proc. APSIPA ASC, 2012.

[10] K.Yoshino, S.Mori, and T.Kawahara.
Language modeling for spoken dialogue system based on sentence transformation and filtering using predicate-argument structures.
In Proc. APSIPA ASC, 2012.

[11] Y.Akita, M.Watanabe, and T.Kawahara.
Automatic transcription of lecture speech using language model based on speaking-style transformation of proceeding texts.
In Proc. INTERSPEECH, 2012.

[12] R.Gomez and T.Kawahara.
Dereverberation based on wavelet packet filtering for robust automatic speech recognition.
In Proc. INTERSPEECH, 2012.

[13] T.Kawahara, T.Iwatate, and K.Takanashi.
Prediction of turn-taking by combining prosodic and eye-gaze information in poster conversations.
In Proc. INTERSPEECH, 2012.

[14] T.Kawahara, T.Iwatate, T.Tsuchiya, and K.Takanashi.
Can we predict who in the audience will ask what kind of questions with their feedback behaviors in poster conversation?
In Proc. Interdisciplinary Workshop on Feedback Behaviors in Dialog, pp.35--38, 2012.

[15] T.Kawahara.
Transcription system using automatic speech recognition for the Japanese Parliament (Diet).
In Proc. AAAI/IAAI, pp.2224--2228, 2012.

[16] T.Kawahara.
Multi-modal sensing and analysis of poster conversations toward smart posterboard.
In Proc. SIGdial Meeting Discourse & Dialogue, pp.1--9 (keynote speech), 2012.

[17] R.Miyazaki, H.Saruwatari, K.Shikano, and K.Kondo.
Musical-noise-free speech enhancement based on iterative Wiener filtering.
In Proc. IEEE Int'l Sympo. Signal Processing & Information Technology (ISSPIT), 2012.

[18]M.Itoi, R.Miyazaki, T.Toda, H.Saruwatari, and K.Shikano.
Blind speech extraction for non-audible murmur speech with speaker's movement noise.
In Proc. IEEE Int'l Sympo. Signal Processing & Information Technology (ISSPIT), 2012.

[19]Y.Takahashi, R.Miyazaki, H.Saruwatari, and K.Kondo.
Theoretical analysis of musical noise in nonlinear noise reduction based on higher-order statistics.
In Proc. APSIPA ASC, 2012.

[20]K.Nishimura, H.Kawanami, H.Saruwatari, and K.Shikano.
Response generation based on statistical machine translation for speech-oriented guidance system.
In Proc. APSIPA ASC, 2012.

[21]S.Kanehara, H.Saruwatari, R.Miyazaki, K.Shikano, and K.Kondo.
Comparative study on various noise reduction methods with decision-directed a priori SNR estimator via higher-order statistics.
In Proc. APSIPA ASC, 2012.

[22]Y.Onuma, N. Kamado, H.Saruwatari, and K.Shikano.
Real-time semi-blind speech extraction with speaker direction tracking on Kinect.
In Proc. APSIPA ASC, 2012.

[23]S.Hara, H.Kawanami, H.Saruwatari, and K.Shikano.
Development of a toolkit handling multiple speech-oriented guidance agents for mobile applications.
In Proc. Int'l Workshop Spoken Dialog Systems (IWSDS), pp.195-200, 2012.

[24]H.Majima, R.Torres, H.Kawanami, S.Hara, T.Matsui, and H.Saruwatari, and K.Shikano.
Evaluation of invalid input discrimination using BOW for speech-oriented guidance system.
In Proc. Int'l Workshop Spoken Dialog Systems (IWSDS), pp.339-347, 2012.

[25]H.Majima, R.Torres, Y.Fujita, H.Kawanami, T.Matsui, H.Saruwatari, K.Shikano.
Spoken inquiry discrimination using bag-of-words for speech-oriented guidance system.
In Proc. INTERSPEECH, 2012.

[26]R.Miyazaki, H.Saruwatari, K.Shikano, and K.Kondo.

Musical-noise-free blind speech extraction using ICA-based noise estimation with channel selection.
In Proc. Int'l Workshop Acoustic Signal Enhancement (IWAENC), 2012.

[27]S.Kanehara, H.Saruwatari, R.Miyazaki, K.Shikano, and K.Kondo.
Theoretical analysis of musical noise generation in noise reduction methods with decision-directed a priori SNR estimator.
In Proc. Int'l Workshop Acoustic Signal Enhancement (IWAENC), 2012.

[28]H.Saruwatari, R.Wakisaka, K.Shikano, and F.Mustiere, L.Thibault, H.Najaf-Zadeh, and M.Bouchard.
Sound-localization-preserved binaural MMSE STSA estimator with explicit and implicit binaural cues.
In Proc. EUSIPCO, pp.310-314, 2012.

[29]R.Miyazaki, H.Saruwatari, K.Shikano, and K.Kondo.
Musical-noise-free blind speech extraction using ICA-based noise estimation and iterative spectral subtraction.
In Proc. Int'l Conf. Information Science, Signal Processing and their Applications (ISSPA), pp.322-327, 2012.

[30]T.Tung and T.Matsuyama.
Invariant Surface-Based Shape Descriptor for Dynamic Surface Encoding,
In Proc. ACCV (LNCS 7724), 201

[31]T.Tung, R.Gomez, T.Kawahara, and T.Matsuyama.
Group dynamics and multimodal interaction modeling using a smart digital signage.
In Proc. ECCV Workshop Video Event Categorization, Tagging and Retrieval (LNCS 7583), pp.362--371, 2012.

[32]Y.Murawaki and S.Kurohashi.
Semi-Supervised Noun Compound Analysis with Edge and Span Features.
In Proc. COLING, pp. 1915-1931, 2012.

[33]J.Harashima and S.Kurohashi.
Flexible Japanese sentence compression by relaxing unit constraints.
In Proc. COLING, pp. 1097-1112, 2012.

[34]M.Hangyo, D.Kawahara, and S.Kurohashi.
A Diverse Document Leads Corpus Annotated with Semantic Relations.
In Proc. Pacific Asia Conf. Language, Information, & Computation (PACLIC), 2012.