戦略的創造研究推進事業　CREST

研究領域
「共生社会に向けた人間調和型情報技術の構築」

研究課題
「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」

研究終了報告書

研究期間　平成21年10月～平成27年3月

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）

§1 研究実施の概要

(1) 実施概要

複数の人間による知的活動のマルチモーダルなインタラクションを長時間収録した音声・映像に対して、人間の直感に基づいて有用箇所を効率的に視覚化・提示できる人間調和型情報基盤を構築した。

具体的には、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・実装を行った。ポスター発表は今でも紙を用いる場合が多く、センサを備えたインタラクション環境は世界的にも例がない。また、このように長時間の複数人による自然な振る舞いを対象として、マルチモーダルな信号処理を行った例もほとんどない。

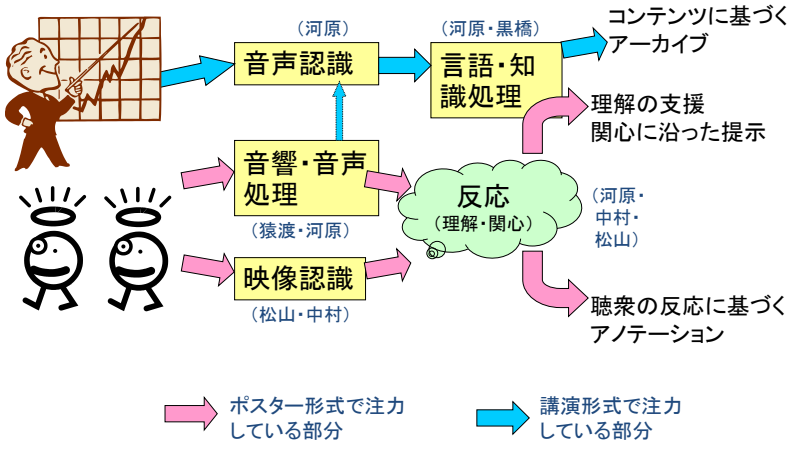
本研究では、音声・音響・映像に関する研究者が結集して、視線や相槌・質問などの聴衆の反応をセンシングすることにより、興味・理解度を推定する枠組みを提案した。実際に、音響と映像の各々の要素処理を密に統合することにより、話者区間検出や興味・理解度推定において相乗的な効果が得られることを示した。

構築したスマートポスターボードシステムは、国際会議を含む様々な学会等において、デモ展示を行った。視線や相槌などの情報を含めて、センシングから機械学習のレベルまで包括的に扱った会話のマルチモーダルな分析は他にほとんど事例がなく、今後、人間どうしのように自然な会話を行えるロボットなどに応用できると期待できる。

また、講演形式のセミナーに対する字幕付与を目標として、話し言葉の音声認識の高度化にも取り組んだ。本研究実施前から取り組んできた衆議院の会議録作成システムについては、2011 年度からすべての会議を対象として本格運用されるようになった。その後も継続的に維持・改良を行った結果、全会議の平均で 90%を上回る認識精度が得られている。

2012 年度からは、京都大学 OCW(OpenCourseWare)で配信されている講演動画への字幕付与の研究開発を行った。収録条件のよい講演については 80%程度の認識率が得られ、これに基づいて作成した字幕を提供している。さらに最近は、放送大学の講義番組に対する字幕付与への展開を進めている。

本研究の処理の概要



(2)　顕著な成果

＜優れた基礎研究としての成果＞

1. 会話のマルチモーダルなセンシングと分析

複数名の聴衆からなるポスターセッションの会話を多様なセンサで多数収録し、マルチモーダルな分析を行った。視線と発話（ターンテイキング）の関係や、視線や相槌と興味・理解度の関係を分析し、実際のセンサデータを用いた機械学習を行い、マルチモーダル統合の意義・効果を実証した。これらの成果については、研究代表者の河原が複数の国際会議で基調講演を行った他、IEEE Transactions on Human-Machine Systems にも（河原と松山らの共著の）論文の掲載が決定した。またこれに関連する研究に対して、情報処理学会論文賞を受賞した。

2. 頑健な音響処理

マイクロフォンアレイを用いた音声分離や発話区間検出の高精度化を目指して、高次統計量を用いた聴感品質定量化やパラメータ最適化アルゴリズムの提案を行い、世界で初めて「一切の聴感品質変化を起こさない」雑音抑圧法を開発した。本研究成果は、IEEE Transactions on Audio, Speech and Language Processing に複数の論文が掲載され、複数の賞を受賞した。

＜科学技術イノベーションに大きく寄与する成果＞

1. スマートポスターボードシステム

ポスターセッションにおいて、発表者や聴衆に一切のセンサを装着してもらうことなく、音声・映像を収録し、話者区間（誰がいつ発話したか）検出と、視線（誰がいつどこを注視していたか）検出を行う処理系を構成した。これにより、長時間会話の効率的なブラウジング・検索を可能にした。このシステムについては、国際会議 IEEE-ICASSP（2012 年 3 月）や日本音響学会（2014 年 3 月）、情報処理学会（2015 年 3 月予定）などで、（河原と猿渡らの共同で）デモ展示を行っている。

2. 議会の会議録作成のための音声認識システム

衆議院の会議録作成のための音声認識に関して継続的に研究開発を進めてきたが、2011 年度からすべての本会議・委員会を対象として運用が行われるようになった。その後も維持・改良を行った結果、全会議の平均で 90%を上回る認識精度が得られている。本研究開発に対して、文部科学大臣表彰（科学技術賞）や情報処理学会喜安記念業績賞などを授与された。

3. 講演の字幕付与のための音声認識システム

講演動画に対して字幕付与を行うための音声認識システムの研究開発を行った。実際に、京都大学 OCW(OpenCourseWare)で一般向けに配信されている数十の講演に対して、字幕付与・配信を行った。現在、放送大学の講義に対する字幕付与に展開している。

§2 研究実施体制

(1) 研究チームの体制について

① 京大グループ

研究参加者

氏名	所属	役職	参加時期
河原 達也	京都大学 学術情報メディアセンター	教授	H21.10～H27.3
中村 裕一	同上	教授	H21.10～H27.3
黒橋 禎夫	京都大学 情報学研究科	教授	H21.10～H25.9
松山 隆司	同上	教授	H21.10～H27.3
秋田 祐哉	京都大学 学術情報メディアセンター	助教	H21.10～H27.3
近藤 一晃	同上	助教	H21.10～H27.3
小泉 敬寛	同上	助教	H21.10～H27.3
柴田 知秀	京都大学 情報学研究科	助教	H21.10～H25.9
河原 大輔	同上	准教授	H23.4～H25.9
Tung Tony	京都大学 学術情報メディアセンター	特定助教	H22.2～H26.9
吉本 廣雅	同上	特定研究員	H22.4～H27.3
村脇 有吾	同上	特定助教	H23.4～H25.9
三村 正人	同上	研究員	H22.4～H27.3
Welly Naptali	同上	研究員	H23.5～H24.3
高梨 克也	同上	研究員	H25.4～H27.3
若林 祐幸	同上	技術補佐員	H24.10～H26.9

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（H21.10～H26.3）

研究参加者

氏名	所属	役職	参加時期
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	H21.10～26.3
猿渡　洋	同上	准教授	H21.10～26.3
原　直	同上	助教	H24.4～25.3
川波　弘道	同上	助教	H25.10～26.3

研究項目

・セミナー・会議のための音響・音声処理

③ 東大グループ（H26.4～H27.3）

研究参加者

氏名	所属	役職	参加時期
猿渡　洋	東京大学 情報理工学系研究科	教授	H26.4～H27.3
小山　翔一	同上	助教	H26.4～H27.3

研究項目

・セミナー・会議のための音響・音声処理

(2)　国内外の研究者や産業界等との連携によるネットワーク形成の状況について

・ 武田チーム・後藤チーム・徳田チームと合同で、2012年4月に京都大学において国際シンポジウム（http://www.ar.media.kyoto-u.ac.jp/crest/sympo12/）を開催し、海外の関連研究者との情報交換・ネットワーク形成につとめた。

§3 研究実施内容及び成果

本研究のメインであるスマートポスターボードのためのマルチモーダル処理に重点をおいて、トピック毎に研究実施内容と得られた成果（原著論文発表の番号を参照）を述べる。

3.1 ポスター形式における支援環境（スマートポスターボード） [32, 60]

3.1.1 研究のコンセプト・独自性

マルチモーダルな信号・情報処理に関する研究は従来、人間型ロボットを含むヒューマンマシンインターフェースの高度化を主な目標として行われてきた。一方、画像処理や音声処理が高度になり、上記のようなインターフェースを意識しない人間の自然なふるまいも扱えるようになり、いわゆるアンビエントなシステムを目指した研究開発も可能になっている。実際に、ミーティングや自由会話などの人間どうしの会話を対象とした研究も行われている。

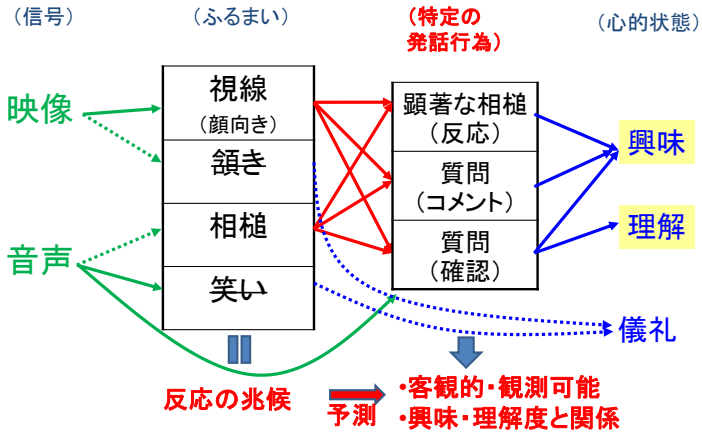
これに対して本プロジェクトでは、ポスターセッションにおける会話（=ポスター会話）に焦点をおいた研究を進めてきた。ポスターセッションは、学会やオープンラボなどで一般的になっているが、未だに情報通信技術(ICT)の導入がほとんどなされておらず、ICT 分野の会議でも紙のポスターを用いることが多い。一部液晶ディスプレイや携帯プロジェクタを用いる場合もあるが、センサを備えた環境は世界的にも前例がないと思われる。講演や講義の映像・音声が収録・配信されることが一般的になっているのに対して、ポスターセッションを収録して分析した研究も皆無に近い。

ポスター会話は、講演と会議の中間的な形態と捉えることができる。すなわち、発表者が自身の研究内容について少人数の聴衆に説明する一方、聴衆の側も相槌や頷きなどでリアルタイムにフィードバックし、時折質問やコメントも行う。また会議と違って、参加者は立っており、動くこともできるので、マルチモーダルなインタラクションを行うことが多い。さらに、ポスター会話を扱う利点としては、話題や他の参加者に対する親近性を制御しながら、（研究者を集めてくれば）自然でリアルなデータを収集することが非常に容易であることが挙げられる。

本研究の目標は、人間どうしのインタラクションの信号レベルのセンシングとより高いレベルの解析である。人間型ロボットを含むヒューマンマシンインターフェースと比較して、長時間にわたる自然な振る舞いを扱う点が最大の違いである。認識のタスクとしては、人物・視線・話者・発話区間などの検出がある。これらは会話アーカイブに対する新たなインデキシングの枠組みを提供する。例えば、自身あるいは同僚のポスターセッションが終わった後で、どのくらいの聴衆がやって来て、どのような質疑・コメントが行われたかを振り返ることができるようになる。

さらに、どの部分に興味を持ってもらえたか、どこがわかりにくいところであったか、といった解析も研究する。動画投稿サイトの事例からもわかるように、我々は他の人が興味を持ったものを視聴したくなるのが普通であるので、このようなアノテーションは有用であるが、アノテーションの基準や評価を含めて、これらを明確に定義するのは非常に困難である。そこで、これらの心的状態に関係し、客観的に観測できる発話行為に着目する。具体的には、聴衆による質問と特定のパターンの相槌に着目する。さらに、質問を確認質問と踏み込み質問に分類する。マルチモーダルな振る舞いからこれらの発話行為を予測することで、興味・理解度の推定を近似する。提案する枠組みを図1に示す。

図1　マルチモーダルなセンシング・解析の枠組み



3.1.2 スマートポスターボードを用いた会話データの収録

本研究では、ポスター会話における音声・映像と振る舞いなどのマルチモーダルな情報を収録するための環境の構築を進めてきた。具体的には、ポスターボードを大型液晶ディスプレイで構成し、これに多様なセンサを設置することで、ポスター会話をセンシングするシステム（=スマートポスターボード）を構築した。スマートポスターボードの概観を図2に示す。

音声に関しては、ポスターボードの上に設置するマイクロフォンアレイを設計した。映像に関しては、参加者全員とポスターをカバーできるように、6～8 個のカメラをポスターボードに設置した。また、Kinect センサも設置した。なお簡易版は Kinect センサのみを用いる。

ただしコーパス構築の上では、正確な情報を取得する必要がある。そのために、各参加者にワイヤレスのヘッドセットマイクを装着してもらうとともに、様々なセンサを着用してもらった。当初はモーションキャプチャシステムや視線計測装置を使用したが、直近は磁気センサを使用している。

図2　スマートポスターボード

上記の環境を用いて、これまで5ヶ年度にわたって合計 43 セッションのポスター会話を収集してきた。ただし、いくつかのセンサデータが欠損したものも含まれる。

各セッションにおいては、1名の発表者(Aと表記)が自身の研究に関する発表を、2名の聴衆(B, C と表記)に対して行う。聴衆は、発表者についても研究内容についても初めて接する設定となっている。セッションの長さは制御しているわけではないが、おおむね 20～30 分程度である。

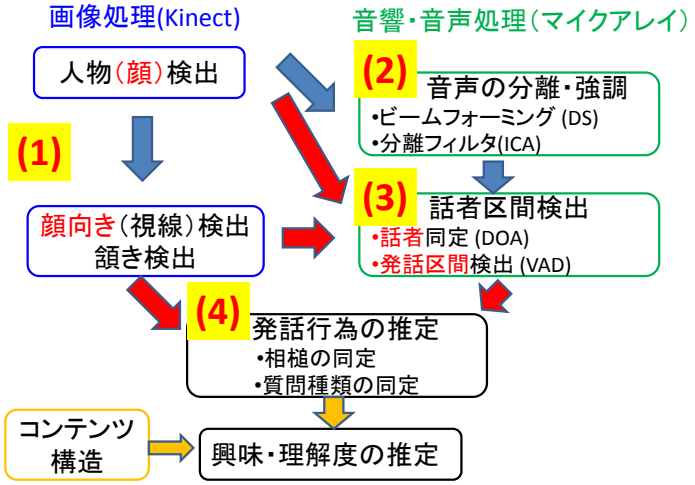
音声データは、ヘッドセットマイクで収録されたものをポーズで区切られた発話単位(IPU)に分割し、時間と話者ラベルを付与した上で、『日本語話し言葉コーパス』(CSJ)と同じ基準で書き起こしを行った。ただし、フィラー以外に相槌と笑いに対してもアノテーションを行った。視線情報は、視線計測装置とモーションキャプチャシステム、または磁気センサのデータを用いて、視線ベクトルと他の参加者やポスターの位置との交差判定に基づいてアノテーションを行った。

3.1.3 マルチモーダルな振る舞い(視線・発話)の検出

音声や映像の信号から、各会話参加者の振る舞いを検出する処理については、順次研究開発を進めてきた。ここでの目標は、参加者にマイクなどの装置を一切装着してもらうことなく、ポスターボード等に設置したセンサだけで処理を行うことである。具体的には、マイクロフォンアレイで収録される音声信号、及び複数台のカメラ(Kinect センサ)で得られる映像信号・距離情報を用いる。

各処理の高度化と適用を行なうとともに、これらのマルチモーダルな情報を統合する方法を検討してきた。全体の処理の流れを図3に示す。まず聴衆の人物(顔)を検出し、各人の顔向き(視線)を検出・追跡する。人物の位置情報は、音響処理、具体的には音声強調と話者同定において利用される。また、視線情報は発話区間検出や質問種類の同定に用いる。

図3　マルチモーダル統合による
ふるまい検出の処理の流れ



(1)　Kinect センサによる視線(頭部方向)検出 [41]

本研究ではKinectセンサを用いて、ポスター会話に適した実用的な視線推定を実現した。なお、画像の解像度及び眼鏡等の影響により眼球そのものを安定に撮影することは困難であるため、視線方向を頭部方向で代用する。視線と頭部方向のずれは平均 10 度程度で、ポスターを注視する状況ではさらに小さくなる傾向がある。処理は以下の手順で行っている。この処理は、GPU を使うことで、オンライン・リアルタイムに実行可能である。

① 顔検出

Kinect センサで撮影したカラー画像及び距離画像から、Haar-like 特徴を利用した物体認識法を用いて正面顔探索を行う。同時に複数人を処理することが可能である。

② 頭部モデル獲得

顔検出結果に従って、距離画像から頭部の 3 次元形状を、カラー画像からその色情報を計算する。計算結果はポリゴンとテクスチャ情報に変換し、頭部モデルとする。

③ 頭部追跡

頭部方向の推定を、画像への頭部モデルのフィッティングとして行う。具体的には、頭部を剛体とみなし、頭部の 3 次元位置と姿勢を表す 6 変数をパーティクルフィルタによる追跡処理で逐次計算する。

④ 注視対象特定

頭部追跡処理で得られた 6 変数から、3 次元空間で視線に対応する半直線を求める。この半直線と、ポスターボードや他の参加者との交差判定を行うことで、注視対象を決定する。

(2)　マイクロフォンアレイによる音声の分離・強調 [8, 49, 66, 88]

音声の分離と強調は、ブラインド空間的サブトラクションアレイ(BSSA)によって実現する。これは、マイクロフォンアレイで得られる信号に対して、遅延加算(DS)型ビームフォーミングを行うとともに、独立成分分析(ICA)に基づいて各会話参加者の音声と背景雑音を分離し、目的信号以外の抑圧を行うものである。ポスター会話の設定では、発表者・聴衆・背景雑音の3つの成分への分離を行う(聴衆間の分離は行われない)。その際に、画像処理によって得られる各参加者の位置情報を用いることで、ICA のフィルタ計算の高速化を実現している。この処理を逐次的に行うことで、参加者が移動しても追随できるようにしている。

19 チャネルのマイクロフォンアレイを用いる場合は、高い品質の音声強調ができるが、リアルタイムには処理できない。Kinect センサ内蔵の複数のマイクロフォンを用いる場合は、音質は低下するが、リアルタイム処理が可能である。

(3)　音響情報と画像情報(話者位置と視線情報)を統合した話者区間検出 [17, 15, 57]

話者区間検出は、「いつ誰が発話したか」を検出する処理で、話者同定と発話区間検出(VAD)の 2 つの要素からなる。そのために、マイクロフォンアレイで得られる音響情報(音のパワーと位相情報)に加えて、画像から得られる各参加者の位置情報を利用する。マルチモーダルな発話区間検出として、口唇の動きを用いることも考えられるが、解像度の高い正面画像が必要なため、ポスタ

一会話においては現実的でない。

まず、マイクロフォンアレイを用いた音源(DOA)の到来方向推定の代表的な手法であるMUSIC法を用いる。MUSIC 法は、観測信号の部分空間の直交性に基づいて、同時に複数の音源をリアルタイムに推定することができ、各時刻・各方向に関して、そこに音源が存在する尤度を求めることができる。

従来のベースライン手法では、この尤度の極大値を探索し、しきい値以上となるものを音源とみなす。このときに、画像処理による顔検出で得られる参加者の位置情報を利用する。すなわち、尤度がしきい値以上であり、かつこの方向が参加者の推定位置からしきい値以内である場合に、発話がなされたと判定する。本研究では、確率的な統合手法を定式化した。これは、画像情報により得られる参加者の位置情報に関して、推定位置を平均、信頼度を分散とする正規分布に基づいて尤度を算出し、上記の尤度と統合するものである。

さらに、視線情報の利用も検討した。視線配布は発話権交替(ターンテイキング)と関係があることが知られている。本研究ではまず、ポスター会話においてその関係を調べた。図4から、誰が発話権を取得するかは参与者相互の視線と相関があるが、いつ発話権交替が生じるかは主に発表者の視線で決まることが示された。そして、各参与者の視線の情報から、発話権交替が起こるかを70%程度の精度で予測することができることを示した。これに基づいて、話者区間検出において、視線の情報を音響情報と確率的な枠組みで統合する方法を定式化した。

話者区間検出の結果を表1に示す。発表者はマイクに近く、発話が多いため、90%に近い検出精度(誤り率 10%)が得られるのに対して、聴衆の発話区間検出は容易でない。単純なベースライン手法では 75%程度であった。これに対して、画像による位置情報を用いることで改善が得られた。さらに、視線の情報を導入することで改善が得られ、最終的には 80%程度の精度が得られた。

図4 視線配布頻度と発話者交替(聴衆による発話権取得)との関係

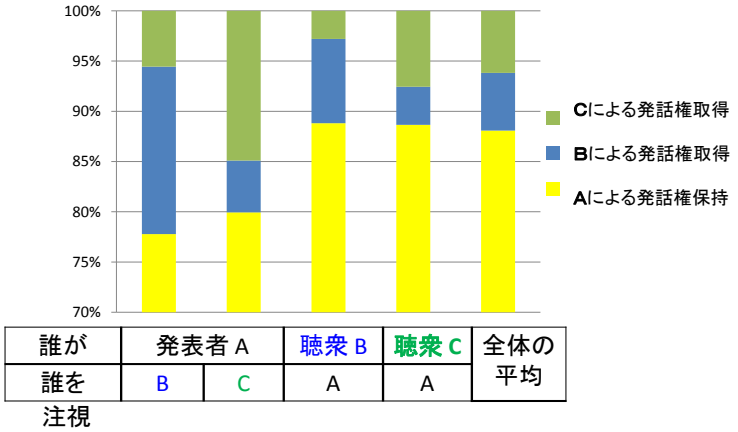


表1　話者区間検出における画像情報（話者位置と視線）統合の効果

	発表者の平均誤り率	聴衆の平均誤り率
ベースライン(MUSIC)	14.7% (21.8%)	25.8% (46.9%)
＋画像位置情報の統合	19.9% (20.1%)	21.2% (34.3%)
＋視線情報の統合	11.7% (18.4%)	18.9% (31.8%)

() 内は 0dB の雑音付加時

(4)　相槌と笑い声の検出 [81, 107]

相槌と笑い声は、聴衆の非言語的な音声による反応として重要であり、Podcast などの会話コンテンツを対象に、検出方法の研究を行ってきた。これは、GMM によるモデル化と BIC によるセグメンテーションを組み合わせた方法である。

ただし相槌は、通常の音声と周波数特徴が類似しているため、韻律的特徴を併用する。聴衆の相槌のうち、特に重要な反応と考えられるのは、「へー」「あー」「ふーん」などの非語彙的で引き延ばし型のものであるので、持続長が長く、基本周波数や周波数包絡が一定時間平坦なものを抽出する。

笑い声は GMM による単純な方法で、再現率・適合率とも約 70%の検出が可能であるが、相槌は明瞭でないものが多いので、再現率約 30%・適合率約 80%となっている。ただし、聴衆の明白な反応を検出できれば十分と考えている。

これにより、聴衆の興味を捉えられることを示している。

3.1.4 マルチモーダルな振る舞いからの発話行為の推定による興味・理解度の推定 [34]

興味・理解度のアノテーションを行う最も自然な方法は、ポスターセッション終了後に聴衆の各人に、各々のスライド話題単位（＝ポスター中の各スライドに対応）に対する興味と理解の評定を行ってもらうことである。しかしながら、このようなアンケート調査を大規模に行うことはあまり現実的でないし、既に収録済みのセッションに行うことは不可能である。またこのような評定は主観的で、その信頼性を評価することも難しい。

そこで本研究では、興味・理解度に関係が深いと考えられ、客観的に観測可能な発話行為に着目した。これまでに、「へー」「あー」「ふーん」といった非語彙的・引き延ばし型で韻律的にも顕著な特徴を持つ相槌（=顕著な相槌）が聴衆の興味と関係があることを明らかにしている。 [107]

また経験的に、聴衆の質問の生起は興味と関係があると考えられる。すなわち、聴衆は発表に引きつけられるほど、より多くの質問をするものである。また、質問の種類を調べることで、理解度を推測することもできる。例えば、既に説明されたことを質問しているなら、理解が困難であったことを示唆している。

本研究では、質問を確認質問と踏み込み質問に分類した。確認質問は、現在の説明の理解が正しいか確認するために行うもので、「はい／いいえ」のいずれかで答えることができる。これに対して踏み込み質問は、発表者の説明に含まれていなかったことに関して尋ねるもので、「はい／いいえ」のみで答えられるものでなく、何らかの補足説明が必要になる。踏み込み質問は、表層的には質問の形式をとっているが、実質的にコメントに近い場合もある。

2012 年度に収録した 4 つのセッションについては、終了後に聴衆の各人に各スライド話題単位に対する興味と理解の度合いを評定してもらった。そこで、このデータを用いて、評定と質問との関係を調べた。

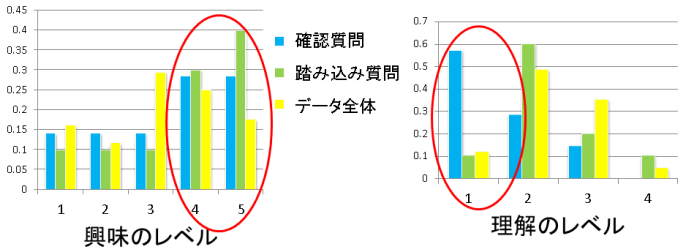
図5に、2種類の質問(確認質問・踏み込み質問)の生起毎、及び全話題セグメント(＝聴衆の各人毎に定義されるスライド話題単位毎の対話セグメント)の興味・理解度の分布を示す。興味度については、1(低い)から 5(高い)の 5 段階で評定してもらい、理解度については、1(低い)から 4(高い)の 4 段階で評定してもらっている。左のグラフから、質問の種類に関わらず、質問が生起している場合には全般に興味が高い(4 か 5)ことがわかる。また右のグラフから、確認質問の大多数(86%)が理解度が低い(1 か 2)ことと相関があることがわかる。

この分析結果と顕著な相槌に関する先行研究 [107]を踏まえて、分析対象の全話題セグメントに対して、以下のアノテーションの枠組みを採用した。

- 興味が高い ← (種類に関わらず)質問もしくは顕著な相槌が生起している
- 理解度が低い ← 確認質問が生起している

これらのアノテーションは各話題セグメントに対して、聴衆 2 名の各人について行った。これらの心的状態の検出は、ポスター会話を後で振り返る際に有用であると考えられる。

図5 質問の生起・種類と興味・理解度の関係



質問が多い→興味度高い　　確認質問→理解度低い

聴衆の興味・理解度の推定を、関係する発話行為の予測を行う問題として定式化した。すなわち、興味の推定は質問と顕著な相槌の生起の予測に、理解度の推定は質問タイプの分類に帰着させた。この予測を、当該の発話行為が(話題セグメントの終わり頃に)実際に生起する前に、聴衆のマルチモーダルな振る舞いを元に行う。これにより、そのような発話行為が実際に生起しなくても、聴衆の心的状態の推定が可能になる。

各聴衆のマルチモーダルな振る舞いとして、相槌と視線配布に着目した。相槌については、発表者の発話で正規化した平均頻度を求めた。発表者に対する視線配布については、発表者の発話で正規化した出現頻度と継続時間割合を求めた。表2に質問の種類毎のこれらの分布を示す。これから、(1) 質問を行う際には相槌が有意に増える、(2) 踏み込み質問を行う際には発表者への視線配布が増える、(3) 確認質問を行う際には発表者への視線配布は少なくポスターを注視し

ている、といった傾向がわかる。

これらの特徴を用いて識別のための機械学習を行った。ここでは、ナイーブベイズ分類器を用いた。これは、学習データが少なく、各特徴量の重みなどのパラメータを推定することが困難であるためである。また、個々の確率を計算するのには、ヒストグラム量子化を用いた。これは、特徴量の値を量子化ビンに割り当てるもので、確率密度関数を仮定しないためモデルパラメータの推定を必要としない。特徴量の分布ヒストグラムを単純に 3 ないし 4 に分割して量子化ビンを設定する。その上で、各ビンの相対的な出現頻度を確率値に変換する。

本実験では 2009 年度と 2011 年度に収録したものを含めて合計 10 セッションを用いた。この 10 セッションには計 58 個のスライド話題単位があった。各セッションに 2 名の聴衆がいるので、興味・理解度を推定すべきスロット(= 話題セグメント)が合計 116 個あることになる。評価実験は、セッション単位の leave-one-out クロスバリデーションにより行った。

(1) 質問・顕著な相槌の生起の予測:興味度の推定

まず、各話題セグメントにおける聴衆の興味度を推定する実験を行った。これは聴衆が質問ないし顕著な相槌を生成するかを予測する問題に帰着される。すなわち、当該の発話行為を行った聴衆は、その話題セグメントに「興味を持った」とみなす。

種々の特徴量に対する正解率を表3(左列)に示す。なお、すべての話題セグメントに「興味を持った」とした場合(chance rate)のベースラインは、49.1%である。

相槌と視線の特徴を用いることで、有意に高い正解率が得られ、両者を組み合わせることで70%を上回る結果となった。ただし、視線に関する2つの特徴量(頻度と時間)については一方を外しても結果は変わらなかった。以上、相槌と視線のマルチモーダルな統合効果を確認した。

(2) 質問の種類の同定:理解度の推定

次に、各話題セグメントにおける聴衆の理解度を推定する実験を行った。これは聴衆が質問を行った際に、質問の種類を同定する問題に帰着させる。すなわち、確認質問を行った聴衆は、その話題セグメントの「理解が困難であった」とみなす。

確認質問/踏み込み質問の分類結果を表3(右列)に示す。なお、このタスクでは各質問の出現頻度に基づく(chance rate)ベースラインは、51.3%である。

すべての特徴量が正解率の向上に一定の効果があったが、視線の出現頻度のみで最良の正解率が得られた。

表2 マルチモーダルな振る舞いと質問の生起・種類の関係

	確認質問	踏み込み質問	データ全体
(1) 相槌の頻度	0.53	0.59	0.42
(2) 発表者への視線配布頻度	0.38	1.02	0.64
(3) 発表者への視線配布時間割合	0.05	0.15	0.07

表3　マルチモーダルな振る舞いに基づく質問の生起・種類の予測結果

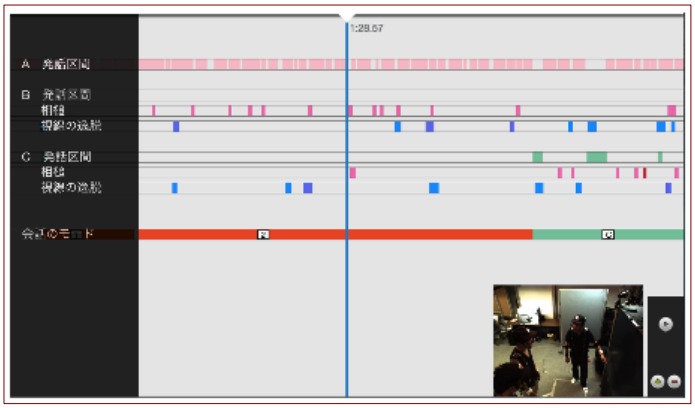
	質問生起の予測精度（興味度推定）	質問種類の同定精度（理解度推定）
ベースライン（chance rate）	49.1%	51.3%
(1) 相槌の頻度	55.2%	56.8%
(2) 発表者への視線配布頻度	61.2%	75.7%
(3) 発表者への視線配布時間割合	57.8%	67.6%
(1)～(3)の組合せ	70.7%	75.7%

3.1.5 ポスター会話のブラウザによる可視化・検索

以上の研究成果に基づいて、ポスター会話のブラウザを作成した。その概観を図6に示す。これは、発表者と聴衆の各人毎に発話区間と特徴的な視線状態を表示するものである。特に、ポスター会話においては、聴衆の質問やコメントが非常に重要であるが、これらはセッション全体の中で時間的にはわずかである。このブラウザにより、聴衆の誰がいつ発話したかを容易に見つけることができる。また、前節の知見から、発表者への視線状態の分布を元に興味・理解度を推測することができる。例えば、これらが皆無であると、興味・理解が低いと推測される。

図6　ポスターセッションブラウザ

- 聴衆の発言を容易に見つけ出し、再生 ← 話者区間検出
- 興味などの反応も推察可能 ← 視線・相槌検出



3.2　講演形式における支援環境（高度な字幕付与）

講演形式のセミナーにおけるリアルタイムの支援と効果的なアーカイブ化を行った。学術講演のような専門性の高い、話し言葉に対する音声認識の高精度化を行い、字幕付与が行えるレベルを目指した。そのために、講演や講師に対して音声認識システムを効果的・効率的に適応する様々な方法を研究した。また、話し言葉は区切りが明確でなく、そのまま書き起こしても読みづらいので、字幕としてふさわしいように整形や文の区分化を行った。さらに、理解が困難と思われる専門用語の自動検出についても検討した。

本研究における最大のチャレンジは、専門性の高い話し言葉に対して、どのように音声認識や自然言語処理を適応するかという点であるので、様々な講演形式のセミナーの収録とコーパス構築から行った。京都大学では、講義内容を学外に対して広く Web で公開する OCW(OpenCourseWare) を推進しており、近年は講演映像の収録と配信を多数行っているので、これを対象とすることにした。

具体的に以下について研究を実施した。

(1)　音声認識 [55]

予稿集やスライドのテキストから単語辞書と言語モデルを適応し、さらに当該講師の音声に音響モデルを適応する方法を研究した。

まず、予稿集を話し言葉に自動変換した上で言語モデルを構成するアプローチを提案し、音声言語関連の講演を対象に平均約 85%の認識精度を実現している。また、京都大学で行われた講義に対しては、当該講師の過去の講義を用いて適応することで、60～80%程度の認識精度を実現している。さらに、京都大学 OCW で公開されている iPS 細胞研究所のシンポジウム講演に対しては、スライドの文字認識結果や関連テキストを用いて適応を行うことで、80～85%程度の認識精度を実現した。

(2)　書き起こしの整形と構造抽出（字幕付与）[46, 86]

話し言葉の書き起こしに対して、統計的機械翻訳のアプローチに基づいて、整形を行ったり、句読点を挿入する方法を研究した。国会審議の書き起こしを用いた評価で必要な整形の 76%、講演の書き起こしについては 50%が自動化できることを示した。

(3)　質問応答及び情報推薦（字幕の高度化）[76]

予稿やスライドなどから専門用語を自動抽出し、その解説を予稿や Web ページなどから取得する方法を研究した。講演スライドに存在する専門用語の 72%を正しく抽出できることを示した。

京都大学OCW(http://ocw.kyoto-u.ac.jp/)で公開されているiPS細胞研究所のシンポジウムのすべての講演と「大震災後を考える」シンポジウムの一部の講演について、実際に字幕を付与し、公式に配信されている（図7参照）。

また昨年度からは、放送大学の講義を対象に音声認識・字幕付与の研究を開始した。講義テキストを用いて単語辞書と言語モデルを適応することで、85%程度の認識率は得られたが、放送大学には膨大な講義コンテンツがあるので、これを効率的にデータベース化して音響モデルの学習

を行い、さらなる高精度化を目指している。

本研究に関する一般へのアウトリーチ活動の一環として、『聴覚障害者のための字幕付与技術』シンポジウムを毎年京都大学で開催している。聴覚障害者・速記者・要約筆記ボランティア・技術者と意見交換を行いながら、音声認識を用いた字幕付与の実演を行っている。河原の講演に対して、リアルタイムで字幕付与を行っている(図8参照)。

(シンポジウムの詳細は http://www.ar.media.kyoto-u.ac.jp/jimaku/参照)

図7　京都大学OCWの講演に対する字幕付与
http://ocw.kyoto-u.ac.jp

図8 『聴覚障害者のための字幕付与技術』シンポジウム
http://www.ar.media.kyoto-u.ac.jp/jimaku/

32

3.3　衆議院での会議録作成支援のための音声認識（本項目は河原のみが関与）[59]

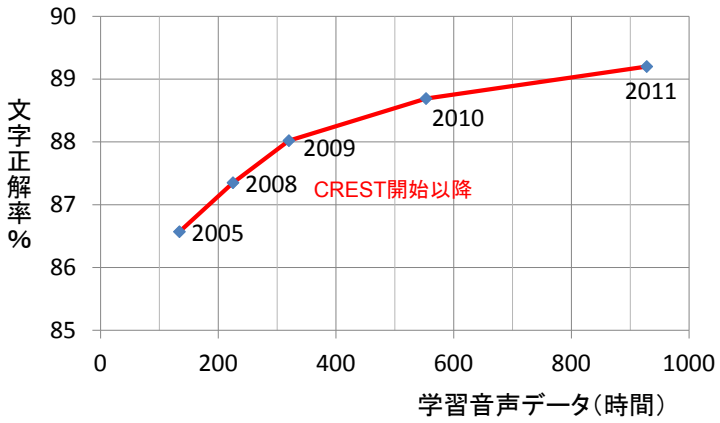
衆議院では、明治 23 年の開設以来用いてきた手書き速記に替えて、音声認識を用いた新たな会議録作成システムを導入した。音声認識の主要モジュールである音響モデル・言語モデル・単語辞書については、河原らの技術が採用されている。本システムを長期的に運用しながら、会議音声に関して世界的にも最大規模のデータを蓄積し、音声認識精度のさらなる改善を図った。

衆議院では年間千時間規模の審議が行われる。これだけの規模の音声を忠実に書き起こすのは事実上不可能である。河原らは、会議録テキストから実際の発言内容を予測する確率的なモデルを考案しており、これに基づいて自動的に高い精度で忠実な発言を復元することができる。この枠組みを継続的に適用・評価した。

本システムは 2010 年度に試験運用を行い、評価と改善を経て、2011 年度から正式運用となった。本研究開始時（2009 年夏）には、225 時間の音声データで音響モデルを構築していた。これに対して、毎年データを追加していき、2011 年末に 928 時間まで増やすことができた。同一の評価セット（2011 年の6会議）に対する文字正解率が図9に示す通り、87.35%から 89.20%に改善した。この規模のデータで報告されている例は世界的にもまれである。絶対値で 1.85%の改善であるが、速記者にとっても実感できるレベルである。

2013 年夏には、さらに 2000 時間規模までデータを増加して、音響モデルの更新を行った。直近 2014 年 1 月～6 月の第 186 回通常国会で運用しているログから集計した実効的な文字正解率は 90.6%に達している。このように、正解率 90%という当初の研究計画書に掲げた目標を達成できている。

図9　衆議院審議の音声認識率の改善



§4 成果発表など

(1) 原著論文発表 (国内(和文)誌 14 件、国際(欧文)誌 105 件)

(2015 年度予定)

[1] K.Yoshino and T.Kawahara.
Conversational system for information navigation based on POMDP with user focus tracking.
Computer Speech and Language, Vol.29, 2015. (DOI: 10.1016/j.csl.2015.01.003)

[2] F.D.Aprilyanti, J.Even, H.Saruwatari, K.Shikano, S.Nakamura, and T.Takatani.
Suppresion of noise and late reverberation based on blind signal extraction and wiener filtering.
Acoustical science and technology, 2015.

[3] 吉本廣雅, 中村裕一.
識別器の特性の学習とユーザの誘導による協調的ジェスチャインタフェース.
ヒューマンインタフェース学会論文誌, 2015.

[4] M.Mimura, S.Sakai, and T.Kawahara.
Deep autoencoders augmented with phone-class feature for reverberant speech recognition.
In Proc. IEEE-ICASSP, 2015.

[5] Y.Akita, Y.Tong, and T.Kawahara.
Language model adaptation for academic lectures using character recognition result of presentation slides.
In Proc. IEEE-ICASSP, 2015.

(2014 年度)

[6] T.Tung, R.Gomez, T.Kawahara, and T.Matsuyama.
Multi-party interaction understanding using smart multimodal digital signage.
IEEE Trans. Human-Machine Systems (THMS), Vol.44, No.5, pp. 625--637, 2014, 2014. (DOI: 10.1109/THMS.2014.2326873)

[7] M.Ablimit, T.Kawahara, and A.Hamdulla.
Lexicon optimization based on discriminative learning for automatic speech recognition of agglutinative language.
Speech Communication, Elsevier, Vol.60, pp.78--87, 2014. (DOI: 10.1016/j.specom.2013.09.011)

[8] R.Miyazaki, H.Saruwatari, S.Nakamura, K.Shikano, K.Kondo, J.Blanchette, and M.Bouchard.
Musical-noise-free blind speech extraction integrating microphone array and iterative spectral subtraction.

Signal Processing, Elsevier, Vol.102, pp.226-239, 2014. (DOI: 10.1016/j.sigpro.2014.03.010)

[9] T. Tung and T. Matsuyama.
Geodesic Mapping for Dynamic Surface Alignment.
IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI) Vol. 36, No.5, pp. 901-913, 2014. (DOI: 10.1109/TPAMI.2013.179)

[10] 小泉敬寛, 小幡佳奈子, 渡辺靖彦, 近藤一晃, 中村裕一.
映像対話型行動支援における頻出パターンに基づいたコミュニケーションの分析.
情報処理学会論文誌, Vol.56, No.3, pp.1068-1079, 2015.

[11] 近藤一晃, 松井研太, 小泉敬寛, 中村裕一.
個人視点映像を対象とした広視野貼り合わせのための画像選択.
電子情報通信学会論文誌, Vol.J98-A, No.1, pp.3-16, 2015.

[12] 吉本廣雅, 中村裕一.
未知剛体の形状と姿勢の実時間同時推定のための Cubistic 表現.
電子情報通信学会論文誌, Vol.J97-D, No.8, pp.1218-1227, 2014.

[13] T.Kawahara, M.Uesato, K.Yoshino, and K.Takanashi.
Toward adaptive generation of backchannels for attentive listening agents.
In Proc. Int'l Workshop Spoken Dialogue Systems (IWSDS), 2015.

[14] K.Yoshino and T.Kawahara.
News navigation system based on proactive dialogue strategy.
In Proc. Int'l Workshop Spoken Dialogue Systems (IWSDS), 2015.

[15] Y.Wakabayashi, K.Inoue, H.Yoshimoto, and T.Kawahara.
Speaker diarization based on audio-visual integration for smart posterboard.
In Proc. APSIPA ASC, 2014.

[16] M.Mimura and T.Kawahara.
Unsupervised speaker adaptation of DNN-HMM by selecting similar speakers for lecture transcription.
In Proc. APSIPA ASC, 2014.

[17] K.Inoue, Y.Wakabayashi, H.Yoshimoto, and T.Kawahara.
Speaker diarization using eye-gaze information in multi-party conversations.
In Proc. INTERSPEECH, 2014.

[18] S.Li, Y.Akita, and T.Kawahara.
Corpus and transcription system of Chinese Lecture Room.
In Proc. Int'l Sympo. Chinese Spoken Language Processing (ISCSLP), 2014.

[19] K.Yoshino and T.Kawahara.
Information navigation system based on POMDP that tracks user focus.
In Proc. SIGdial Meeting Discourse & Dialogue, pp.32--40, 2014.

[20] M.Mimura, S.Sakai, and T.Kawahara.
Exploring deep neural networks and deep autoencoders in reverberant speech recognition.
In Workshop on Hands-free Speech Communication & Microphone Arrays (HSCMA), 2014.

[21] S.Nakai, H.Saruwatari, R.Miyazaki, S.Nakamura, K.Kondo.
Theoretical analysis of biased MMSE short-time spectral amplitude estimator and its extension to musical-noise-free speech enhancement.
In Workshop on Hands-free Speech Communication & Microphone Arrays (HSCMA), 2014.

[22] F.Aprilyanti, H.Saruwatari, K.Shikano, S.Nakamura, T.Takatani.
Optimized joint noise suppression and dereverberation based on blind signal extraction for hands-free speech recognition system.
In Workshop on Hands-free Speech Communication & Microphone Arrays (HSCMA), 2014.

[23] M.Yoshimoto and Y.Nakamura.
Cooperative Gesture Recognition: Learning Characteristics of Classifiers and Navigating User to Ideal Situation.
In Proc. IEEE Int'l Conf. Pattern Recognition Applications and Methods, pp.210-218, 2015.

[24] Y.Nakamura, T.Koizumi, K.Obata, K.Kondo, abd Y.Watanabe.
Behaviors and Communications in Working Support through First Person Vision Communication.
In Proc. IEEE Int'l Conf. Ubiquitous Intelligence and Computing, 2014.

[25] T.Mukasa, S.Nobuhara, T.Tung, and T.Matsuyama.
A 3D Shape Descriptor for Segmentation of Unstructured Meshes into Segment-Wise Coherent Mesh Series.
In Proc. Int'l Conf. 3D Vision (3DV), 2014

[26] T. Tung and T. Matsuyama.
Timing-based Local Descriptor for Dynamic Surfaces.
In Proc. IEEE-CVPR, 2014.

（2013 年度）

[27] 吉野幸一郎, 森信介, 河原達也.
述語項構造を介した文の選択に基づく音声対話用言語モデルの構築.
人工知能学会論文誌, Vol.29, No.1, pp.53--59, 2014. (DOI: 10.1527/tjsai.29.53)

[28] T.Tung and T.Matsuyama.
Invariant shape descriptor for 3D video encoding.

The Visual Computer, Springer, March, 2014. (DOI: 10.1007/s00371-014-0925-6)

[29] T.Mukasa, S.Nobuhara, T.Tung and T.Matsuyama.
Tree-structured mesoscopic surface characterization for kinematic structure estimation from 3D video
IPSJ Transactions on Computer Vision and Applications (CVA), Vol. 6, pp. 12-24, March 2014. (DOI: 10.2197/ipsjtcva.6.12)

[30] 村脇有吾.
階層的複数ラベル文書分類におけるラベル間依存の利用.
自然言語処理, Vol.21, No.1, pp. 41-60, 2014.

[31] 朝倉僚, 宮坂淳介, 近藤一晃, 中村裕一, 秋田純一, 戸田真志, 櫻沢繁.
筋電位計測と画像による姿勢計測を用いたリハビリテーション支援システムの設計.
電子情報通信学会論文誌, Vol.J97-D, No.1, pp.50-61, 2014.

[32] T.Kawahara.
Smart posterboard: Multi-modal sensing and analysis of poster conversations.
In Proc. APSIPA ASC, (plenary overview talk), 2013.

[33] K.Yoshino, S.Mori, and T.Kawahara.
Predicate argument structure analysis using partially annotated corpora.
In Proc. IJCNLP, pp.957--961, 2013.

[34] T.Kawahara, S.Hayashi, and K.Takanashi.
Estimation of interest and comprehension level of audience through multi-modal behaviors in poster conversations.
In Proc. INTERSPEECH, pp.1882--1885, 2013.

[35] K.Yoshino, S.Mori, and T.Kawahara.
Incorporating semantic information to selection of web texts for language model of spoken dialogue system.
In Proc. IEEE-ICASSP, pp.8252--8256, 2013.

[36] S.Nakai, R.Miyazaki, H.Saruwatari, and S.Nakamura.
Theoretical analysis of musical noise generation for blind speech extraction with generalized MMSE short-time spectral amplitude estimator.
In Proc. Intelligent Signal Processing (ISP) Conf., No.4.3, 2013.

[37] H.Saruwatari and R.Miyazaki.
Information-geometric optimization for nonlinear noise reduction systems.
In Proc. Int'l Sympo. Intelligent Signal Processing and Communication Systems (ISPACS), 2013.

[38] R.Miyazaki, H.Saruwatari, S.Nakamura, K.Shikano, and K.Kondo, J.Blanchette, and M.Bouchard.
Toward musical-noise-free blind speech extraction: concept and its applications.

In Proc. APSIPA ASC, 2013.

[39] H.Saruwatari, S.Kanehara, R.Miyazaki, K.Shikano, K.Kondo.
Musical noise analysis for Bayesian minimum mean-square error speech amplitude estimators based on higher-order statistics.
In Proc. INTERSPEECH, pp.441-445, 2013.

[40] H.Yoshimoto and Y.Nakamura.
Cubistic Representation for Real-Time 3D Shape and Pose Estimation of Unknown Rigid Object.
In Proc. ICCV Workshop, pp.522-529, 2013.

[41] H.Yoshimoto and Y.Nakamura.
Free-Angle 3D Head Pose Tracking Based on Online Shape Acquisition.
In Proc. ACPR, pp.798-802, 2013.

[42] T.Tung, R. Gomez, T. Kawahara, and T.Matsuyama.
Multi-party Human-Machine Interaction Using a Smart Multimodal Digital Signage.
In Proc. HCI, LNCS, Vol. 8007, pp.408-415, 2013.

[43] T.Tung and T.Matsuyama.
Intrinsic Characterization of Dynamic Surfaces.
In Proc. IEEE-CVPR, 2013.

[44] Y.Murawaki.
Global Model for Hierarchical Multi-Label Text Classification.
In Proc. IJCNLP, pp. 46-54, 2013.

（2012 年度）

[45] 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.

[46] 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)

[47] 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.

[48] 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.

[49] 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)

[50] 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)

[51] 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)

[52] 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.

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

[54] 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.

[55] 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.

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

[57] 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.

[58] 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.

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

[60] 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.

[61] 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.

[62] 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.

[63] 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.

[64] 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.

[65] 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.

[66] 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.

[67] 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.

[68] 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.

[69] 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.

[70] 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.

[71] 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.

[72] 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.

[73] 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.

[74] T.Tung and T.Matsuyama.
Invariant Surface-Based Shape Descriptor for Dynamic Surface Encoding,
In Proc. ACCV (LNCS 7724), 2012.

[75] 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.

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

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

[78] 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.

（2011 年度）

[79] G.Neubig, M.Mimura, S.Mori, and T.Kawahara.
Bayesian learning of a language model from continuous speech.
IEICE Trans., Vol.E95-D, No.2, pp.614--625, 2012.

[80] 吉野幸一郎, 森信介, 河原達也.
述語項の類似度に基づく情報抽出・推薦を行う音声対話システム.
情報処理学会論文誌, Vol.52, No.12, pp.3386--3397, 2011.

[81] 河原達也, 須見康平, 緒方淳, 後藤真孝.
音声会話コンテンツにおける聴衆の反応に基づく 音響イベントとホットスポットの検出.
情報処理学会論文誌, Vol.52, No.12, pp.3363--3373, 2011.

[82] R.Gomez and T.Kawahara.
Optimized wavelet-based speech enhancement for speech recognition in noisy and reverberant conditions.
Proc. APSIPA ASC, 2011.

[83] M.Mimura and T.Kawahara.
Fast speaker normalization and adaptation based on BIC for meeting speech recognition.
Proc. APSIPA ASC, 2011.

[84] M.Ablimit, T.Kawahara, and A.Hamdulla.
Lexicon optimization for automatic speech recognition based on discriminative learning.
Proc. APSIPA ASC, 2011.

[85] T.Hirayama, Y.Sumi, T.Kawahara, and T.Matsuyama.
Info-concierge: Proactive multi-modal interaction using mind probing.
Proc. APSIPA ASC, 2011.

[86] Y.Akita and T.Kawahara.
Automatic comma insertion of lecture transcripts based on multiple annotations.
Proc. INTERSPEECH, pp.2889--2892, 2011.

[87] R.Gomez and T.Kawahara.
Denoising using optimized wavelet filtering for automatic speech recognition.
Proc. INTERSPEECH, pp.1673--1676, 2011.

[88] T.Inoue, H. Saruwatari, Y.Takahashi, K.Shikano,K.Kondo.
Theoretical analysis of musical noise in generalized spectral subtraction based on higher-order statistics.
IEEE Trans. Audio, Speech & Language Processing, vol.19, no.6, pp.1770-1779, 2011. (DOI:10.1109/TASL.2010.2098871)

[89] H.Saruwatari, Y.Ishikawa, Y.Takahashi, T.Inoue, K.Shikano, K.Kondo.
Musical noise controllable algorithm of channelwise spectral subtraction and adaptive beamforming based on higher-order statistics.
IEEE Trans. Audio, Speech & Language Processing, vol.19, no.6, pp.1457-1466, 2011. (DOI: 10.1109/TASL.2010.2091636)

[90] R.Miyazaki, H.Saruwatari, K.Shikano.
Theoretical analysis of amounts of musical noise and speech distortion in structure-generalized parametric spatial subtraction array.
IEICE Trans. vol.95-A, no.2, pp.586-590, 2012.

[91] R.Wakisaka, H.Saruwatari, K.Shikano, T.Takatani.
Speech prior estimation for generalized minimum mean-square error short-time spectral amplitude estimator.
IEICE Trans. vol.95-A, no.2, pp.591-595, 2012.

[92] K.Kubo, H.Kawanami, H.Saruwatari, K.Shikano.
Unconstrained many-to-many alignment for automatic pronunciation annotation.
Proc. APSIPA ASC, 2011.

[93] Y.Fujita, S.Takeuchi, H.Kawanami, T.Matsui, H.Saruwatari, K.Shikano.
Out-of-task utterance detection based on bag-of-words using automatic speech recognition results.
Proc. APSIPA ASC, 2011.

[94] K.Nishimura, H.Kawanami, H.Saruwatari, K.Shikano.
Investigation of statistical machine translation applied to answer generation for a speech-oriented guidance system.
Proc. APSIPA ASC, 2011.

[95] H.Saruwatari, N.Hirata, T.Hatta, R.Wakisaka, K.Shikano, T.Takatani.
Semi-blind speech extraction for robot using visual information and noise

statistics.
Proc. IEEE ISSPIT, pp.238-243, 2011.

[96] R.Miyazaki, H.Saruwatari, K.Shikano.
Theoretical analysis of musical noise and speech distortion in structure-generalized parametric blind spatial subtraction array.
Proc. INTERSPEECH, pp.341-344, 2011.

[97] R.Wakisaka, H.Saruwatari, K.Shikano, T.Takatani.
Blind speech prior estimaion for generalized minimum mean-square error short-time spectral amplitude estimator.
Proc. of INTERSPEECH, pp.361-364, 2011.

[98] T.Inoue, H.Saruwatari, K.Shikano, K.Kondo.
Theoretical analysis of musical noise in Wiener filtering family via higher-order statistics.
Proc. IEEE-ICASSP, pp.5076-5079, 2011.

[99] 近藤一晃, 西谷英之, 中村裕一.
協調的物体認識のためのマンマシンインタラクション設計.
電子情報通信学会論文誌, Vol. J94-D, No.8, pp. 1206-1215, 2011.

[100] Z.Yu, X.Zhou, Z.Yu, C.Becker, and Y.Nakamura.
Social Interaction Mining in Small Group Discussion Using a Smart Meeting System.
Proc. UIC, Springer-Verlag LNCS, 2011.

[101] Y.Murawaki and S.Kurohashi.
Non-parametric Bayesian Segmentation of Japanese Noun Phrases.
Proc. EMNLP, pp. 605-615, 2011.

（2010 年度）

[102] 三村正人, 秋田祐哉, 河原達也.
統計的言語モデル変換を用いた音響モデルの準教師つき学習.
電子情報通信学会論文誌, Vol.J94-D, No.2, pp.460--468, 2011.

[103] D.Cournapeau, S.Watanabe, A.Nakamura, and T.Kawahara.
Online unsupervised classification with model comparison in the Variational Bayes framework for voice activity detection.
IEEE J. Selected Topics in Signal Processing, Vol.4, No.6, pp.1071--1083, 2010.
(DOI: 10.1109/JSTSP.2010.2080821)

[104] R.Gomez and T.Kawahara.
Robust speech recognition based on dereverberation parameter optimization using acoustic model likelihood.
IEEE Trans. Audio, Speech & Language Processing, Vol.18, No.7, pp.1708--1716,

2010. (DOI: 10.1109/TASL.2010.2052610)

[105] R.Gomez and T.Kawahara.
Optimizing wavelet parameters for dereverberation in automatic speech recognition.
Proc. APSIPA ASC, pp.446--449, 2010.

[106] T.Kawahara.
Automatic transcription of parliamentary meetings and classroom lectures -- a sustainable approach and real system evaluations --.
Proc. Int'l Sympo. Chinese Spoken Language Processing (ISCSLP), pp.1--6 (keynote speech), 2010.

[107] T.Kawahara, K.Sumi, Z.Q.Chang, and K.Takanashi.
Detection of hot spots in poster conversations based on reactive tokens of audience.
Proc. INTERSPEECH, pp.3042--3045, 2010.

[108] T.Kawahara, N.Katsumaru, Y.Akita, and S.Mori.
Classroom note-taking system for hearing impaired students using automatic speech recognition adapted to lectures.
Proc. INTERSPEECH, pp.626--629, 2010.

[109] R.Gomez and T.Kawahara.
An improved wavelet-based dereverberation for robust automatic speech recognition.
Proc. INTERSPEECH, pp.578--581, 2010.

[110] Y.Akita, M.Mimura, G.Neubig, and T.Kawahara.
Semi-automated update of automatic transcription system for the Japanese national congress.
Proc. INTERSPEECH, pp.338--341, 2010.

[111] T.Kawahara, Z.Q.Chang, and K.Takanashi.
Analysis on prosodic features of Japanese reactive tokens in poster conversations.
Proc. Int'l Conf. Speech Prosody, 2010.

[112] T.Inoue, H.Saruwatari, K.Shikano, and K.Kondo.
Theoretical analysis of musical noise in Wiener filter via higher-order statistics.
Proc. of APSIPA ASC, pp.121-124, 2010.

[113] T.Tung and T.Matsuyama.
3D Video Performance Segmentation.
Proc. IEEE-ICIP, pp.25-28, 2010.

[114] T.Tung and T.Matsuyama.
Dynamic Surface Matching by Geodesic Mapping for 3D Animation Transfer.
Proc. IEEE-CVPR, pp.1402-1409, 2010.

[115] P.Huang, T.Tung, S. Nobuhara, A.Hilton, and T. Matsuyama.
Comparison of Skeleton and Non-Skeleton Shape Descriptors for 3D Video.
Proc. Int'l Sympo. 3D Data Processing, Visualization & Transmission (3DPVT), 2010.

[116] K.Kondo, H.Nishitani, and Y.Nakamura
Human-Computer Collaborative Object Recognition for Intelligent Support.
Pacific-Rim Conference on Multimedia (PCM), pp.II-471-482, 2010.

[117] Y.Murawaki and S.Kurohashi
Semantic Classification of Automatically Acquired Nouns using Lexico-Syntactic Clues.
Proc. COLING, Poster Volume, pp. 876-884, 2010.

[118] Y.Murawaki and S.Kurohashi.
Online Japanese Unknown Morpheme Detection using Orthographic Variation.
Proc. Int'l Conf. Language Resources & Evaluation (LREC), pp. 832-839, 2010.

(2009 年度)

[119] 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.

(2) その他の著作物(総説、書籍など)

[1] 河原達也.
音声認識・対話技術の基礎と応用: 最終回 音声対話システムの実際 Siri はどのように成功したか.
日経エレクトロニクス, No. 7-21, pp.86--93, 2014.

[2] 河原達也.
音声認識・対話技術の基礎と応用: 第4回 話し言葉をテキスト化するシステム 会議録の作成や字幕付与への展開.
日経エレクトロニクス, No. 7-7, pp.92--97, 2014.

[3] 河原達也.
音声認識・対話技術の基礎と応用: 第3回 音声認識・対話のアプリケーション 成功の鍵は必然性や自然性.
日経エレクトロニクス, No. 6-23, pp.68--74, 2014.

[4] 河原達也.
音声認識・対話技術の基礎と応用: 第2回 音声認識に新潮流 ビッグデータや DNN を

活用.
日経エレクトロニクス, No. 6-9, pp.82--87, 2014.

[5] 河原達也.
音声認識・対話技術の基礎と応用: 第1回 実用化進む音声認識 システムの構成要素を概観.
日経エレクトロニクス, No. 5-26, pp.88--95, 2014.

[6] H.Saruwatari and R.Miyazaki.
Statistical analysis and evaluation of blind speech extraction algorithms.
In book chapter of Blind Source Separation -Signals and Communication Technology- (Springer) , pp 291-322, May 2014.

[7] 河原達也, 峯松信明.
音声情報処理技術を用いた外国語学習支援.
電子情報通信学会論文誌, Vol.J96-D, No.7, pp.1549--1565, 2013.

[8] 河原達也.
音声認識技術の現状と将来展望.
電気学会誌, Vol.133, No.6, pp.364--367, 2013.

[9] 河原達也.
音声対話システムの進化と淘汰 ―歴史と最近の技術動向―.
人工知能学会誌, Vol.28, No.1, pp.45--51, 2013.

[10] 河原達也.
話し言葉の音声認識の進展 ―議会の会議録作成から講演・講義の字幕付与へ―.
メディア教育研究, Vol.9, No.1, pp.1--8, 2012.

[11] 河原達也.
音声認識技術を用いた講演・講義への字幕付与.
映像情報メディア学会誌, Vol.66, No.8, pp.641--644, 2012.

[12] 河原達也, 秋田祐哉, 三村正人, 堀貴明, 小橋川哲.
2011年度喜安記念業績賞紹介: 議会の会議録作成のための音声認識システムの実用化.
情報処理, Vol.53, No.8, p. 867, 2012.

[13] 河原達也.
国会審議の会議録作成支援のための音声認識システム.
自動認識, Vol.25, No.4, pp.26--29, 2012.

[14] 高橋祐, 宮崎亮一, 猿渡洋.
高次統計量に基づく非線形雑音抑圧処理の数理解析とその応用.
日本音響学会誌, Vol.68, No.11, pp.578-583, 2012.

[15] T. Matsuyama, S. Nobuhara, T. Takai, T. Tung.
3D Video and its Applications, Springer, 2012.
(ISBN: 978-1-4471-4119-8)

[16] 中村裕一.
映像によるライフログ.
情報の科学と技術, Vol.63, No.2, pp.57-62, 2013.

[17] 河原達也.
音声認識技術を用いた講演・講義の字幕配信システム.
日本の速記, No. 870, pp.15--20, (7 月号) 2011.

[18] 鎌土記良, 大沼侑司, 猿渡洋, 鹿野清宏, 高橋祐.
Kinect のマイクロホン・アレーによる音声信号処理.
INTERFACE 1 月号, CQ 出版社, pp.112-116, 2012.

[19] 河原達也.
ロボットのための音声認識.
日本ロボット学会誌, Vol.28, No.3, pp.21--23, 2010.

[20] 河原達也. 日本の国会における音声認識技術を用いた会議録システム (Intersteno 2009 講演要旨). 日本の速記, No. 852, pp.12--17, (11 月号) 2009.

(3) 国際学会発表及び主要な国内学会発表

① 招待講演 (国内会議 15 件、国際会議 10 件)

[1] 猿渡洋.
高次統計量は何を語る？～教師無し学習に基づく自律的な音メディア信号処理～.
情報処理学会研究会 SIG-MUS, 日本大学, 2014 年 5 月 25 日.

[2] T.Tung.
Dynamic Surface Modeling and its Applications.
情報処理学会研究会 SIG-CVIM, 筑波大学, 2014 年 9 月 2 日.

[3] **河原達也.**
スマートポスターボード: ポスター会話のマルチモーダルなセンシングと解析.
人工知能学会研究会 SIG-Challenge, 京都大学, 2014 年 3 月 18 日.

[4] 河原達也.
音声認識の方法論に関する考察—世代交代に向けて—.
情報処理学会研究会 SIG-SLP, 伊豆の国, 2014 年 1 月 31 日.

[5] 河原達也.
音声認識の方法論に関する考察—歴史的変遷と今後の展望—.
情報処理学会研究会 SIG-MUS, お茶の水女子大学, 2013 年 5 月 11 日.

[6] 高梨克也.
マルチモーダルインタラクション分析の基礎と現代的課題.
電子情報通信学会研究会 SP, 名城大学, 2014 年 1 月 24 日.

[7] T.Kawahara.

Smart posterboard: Multi-modal sensing and analysis of poster conversations. ACPR Workshop on Advanced Sensing / Visual Attention and Interaction, 那覇, 2013 年 11 月 5 日.

[8] **T.Kawahara.**
*** Smart posterboard: Multi-modal sensing and analysis of poster conversations.**
APSIPA ASC, (plenary overview talk), 台湾・高雄, 2013 年 10 月 31 日.

[9] T.Kawahara.
Subtitling lecture videos with automatic speech recognition.
Intersteno general conference, ベルギー・ゲント, 2013 年 7 月 16 日.

[10] 猿渡洋.
高次統計量追跡に基づくブラインド信号抽出およびその高品質化.
電子情報通信学会研究会 EA, 東京, 2013 年 6 月 14 日.

[11] **河原達也.**
スマートポスターボード: ポスター発表における場のマルチモーダルなセンシングと認識.
電子情報通信学会研究会 PRMU, 大阪府立大学, 2013 年 2 月 22 日.

[12] 河原達也.
音声対話システムの進化と淘汰.
人工知能学会 SIG-SLUD, 湯河原, 2013 年 2 月 1 日.

[13] **河原達也.**
スマートポスターボード: ポスター会話のマルチモーダルなセンシングと認識.
電子情報通信学会研究会 SP, 天童, 2012 年 7 月 20 日.

[14] **T.Kawahara.**
*** Multi-modal sensing and analysis of poster conversations toward smart posterboard.**
SIGdial Meeting Discourse & Dialogue, (keynote speech), 韓国・ソウル, 2012 年 7 月 5 日.

[15] 猿渡洋.
音声強調処理における高次統計量の利用.
電子情報通信学会研究会 EA, 大阪, 2012 年 5 月 24 日.

[16] T.Kawahara.
New Transcription System using Automatic Speech Recognition (ASR) in the Japanese Parliament (Diet) -- The House of Representatives --.
Intersteno IPRS, フランス・パリ, 2011 年 7 月 14 日.

[17] **河原達也.**
話し言葉の音声認識からコミュニケーションの理解へ -- 国会審議の音声認識からポスター会話の分析へ --.
日本音声学会全国大会 (公開基調講演), 京都大学, 2011 年 9 月 24 日.

[18] 河原達也, 李晃伸.
音声認識ソフトウェア Julius.
人工知能学会全国大会 AI レクチャー, 盛岡, 2011 年 6 月 2 日.

[19] H.Saruwatari.
Recent advances on noise reduction and source separation technology for robot audition.
IEEE/RSJ IROS, 米国・サンフランシスコ, 2011 年 9 月.

[20] T.Kawahara.
Automatic transcription of parliamentary meetings and classroom lectures -- a sustainable approach and real system evaluations --.
Int'l Sympo. Chinese Spoken Language Processing (ISCSLP), **(keynote speech)**, 台湾・台南, 2010 年 12 月 3 日.

[21] T.Kawahara.
Conversation analysis based on reactive tokens in poster sessions.
Workshop on Predictive Models of Human Communication Dynamics, 米国・ロスアンゼルス, 2010 年 8 月 4 日.

[22] 猿渡洋.
音声信号処理における雑音抑圧技術の最新動向.
電子情報通信学会研究会 SP, けいはんな, 2011 年 1 月 27 日.

[23] T.Tung.
3D Video Understanding using a Topology Dictionary.
Dagstuhl Seminar Proc. Computational Video, ドイツ・ダグシュトール, 2010 年 10 月 14 日.

[24] T.Kawahara.
* New perspectives on spoken language understanding: Does machine need to fully understand speech?
IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), イタリア・メラノ, 2009 年 12 月 16 日.

[25] 黒橋禎夫.
構造的言語処理による情報検索基盤の構築と情報分析.
情報アクセスシンポジウム(IAS), 札幌, 2009 年 10 月.

② 口頭発表 （国内会議 165 件、国際会議 41 件）

（国内）

[1] 山口貴史, 井上昂治, 吉野幸一郎, 高梨克也, 河原達也.
傾聴対話における相槌形態と先行発話の統語構造の関係の分析.

人工知能学会研究会資料, SLUD-B403-4, 2015.
[2] 井上昂治, 若林佑幸, 吉本廣雅, 高梨克也, 河原達也.
ポスター会話における音響・視線情報を統合した話者区間及び相槌の検出.
情報処理学会研究報告, SLP-105-9, 2015.
[3] 吉野幸一郎, 河原達也.
ユーザの焦点を用いた POMDP による音声情報案内システム.
人工知能学会研究会資料, SLUD-B402-14, 2014.
[4] 吉野幸一郎, 河原達也.
ユーザの焦点を用いた POMDP による音声情報案内システム.
情報処理学会研究報告, SLP-104-9, 2014.
[5] 三村正人, 坂井信輔, 河原達也.
ディープオートエンコーダと DNN-HMM を用いた残響下音声認識.
情報処理学会研究報告, SLP-102-6, 2014.
[6] Sheng Li, Yuya Akita, and Tatsuya Kawahara.
Classifier-based data selection for lightly-supervised training of acoustic model for lecture transcription.
情報処理学会研究報告, SLP-102-4, 2014.
[7] 井上昂治, 若林佑幸, 吉本廣雅, 高梨克也, 河原達也.
スマートポスターボードにおける視線情報を用いた話者区間検出及び相槌の同定.
情報処理学会全国大会講演論文集, 6P-09, 2015.
[8] 山口貴史, 吉野幸一郎, 高梨克也, 河原達也.
多様な形態の相槌をうつ音声対話システムのための傾聴対話の分析.
情報処理学会全国大会講演論文集, 6P-08, 2015.
[9] 大田健翔, 秋田祐哉, 河原達也.
講演音声認識結果の誤り箇所の復唱入力を用いたノートテイクシステム.
情報処理学会全国大会講演論文集, 5P-06, 2015.
[10] 吉野幸一郎, 河原達也.
ユーザの焦点を用いた POMDP による音声情報案内システム.
情報処理学会全国大会講演論文集, 3D-01, 2015.
[11] 童弋正, 秋田祐哉, 河原達也.
講演スライドの文字認識結果を用いた音声認識の改善.
情報処理学会研究報告, SLP-102-3, 2014.
[12] 井上昂治, 若林佑幸, 吉本廣雅, 河原達也.
多人数会話における視線情報を用いた話者区間検出.
情報処理学会研究報告, SLP-102-1, 2014.
[13] 井上昂治, 若林佑幸, 吉本廣雅, 河原達也.
多人数会話における音響情報と視線情報の確率的統合による話者区間検出.

日本音響学会研究発表会講演論文集, 2-8-4, 秋季 2014.

[14] 吉本廣雅, 中村裕一.
ポスター対話における会話参与者の振舞いの定量化.
情報処理学会全国大会講演論文集, 3D-03, 2015.

[15] 近藤一晃, 小幡佳奈子, 中村裕一.
集合的個人視点映像の自動編集に関する基礎検討 ～屋外グループ活動の効果的な記録・閲覧を目指して～.
HCG シンポジウム, pp.587-589, 2014.

[16] 吉本廣雅, 中村裕一.
ポスター自動発表システムの実現に向けた視線分布に基づくポスター対話の話題推定.
HCG シンポジウム, pp.562-569, 2014.

[17] 保澤圭亮, 吉本廣雅, 近藤一晃, 小泉敬寛, 中村裕一, 古谷栄光.
人間の指差し動作モデルを用いたポインティングシステムの設計と性能予測.
HCG シンポジウム, pp.537-544, 2014.

[18] 小泉敬寛, 小幡佳奈子, 渡辺靖彦, 近藤一晃, 中村裕一.
映像対話型行動支援におけるインタラクションの一貫性に関する考察.
HCG シンポジウム, pp.49-56, 2014.

[19] Du Bang, 近藤一晃, 吉本廣雅, 中村裕一.
Measurement of attention diversity in cooking situation.
映像情報メディア学会年次大会, 2014.

[20] Siyang Yu, Kazuaki Kondo, Hiromasa Yoshimoto, Yuichi Nakamura, Masatake Dantsuji.
Browsing of concentration by capturing learner's behaviors in e-learning.
映像情報メディア学会年次大会, 2014.

[21] 桑原暢弘, 秋田祐哉, 河原達也.
音声認識結果の有用性の自動判定に基づく 講義のリアルタイム字幕付与システム.
音声ドキュメント処理ワークショップ, 2014.

[22] 吉野幸一郎, 河原達也.
ユーザの焦点に適応的な雑談型音声情報案内システム.
人工知能学会研究会資料, SLUD-B303-11, 2014.

[23] 上里美樹, 高梨克也, 河原達也.
傾聴対話における相槌の韻律的特徴の同調傾向の分析.
人工知能学会研究会資料, SLUD-B303-02, 2014.

[24] 河原達也, 林宗一郎, 高梨克也.
ポスター会話における聴衆のマルチモーダルな振る舞いに基づく 興味・理解度の推定.
情報処理学会研究報告, SLP-97-12, 2013.

[25] 三村正人, 河原達也.

CSJ を用いた日本語講演音声認識への DNN-HMM の適用と話者適応の検討.
情報処理学会研究報告, SLP-97-9, 2013.

[26] 吉野幸一郎, 森信介, 河原達也.
述語項構造を介した Web テキストからの文選択に基づく言語モデルの評価.
情報処理学会研究報告, SLP-97-4, 2013.

[27] 吉野幸一郎, 河原達也.
ユーザの焦点に適応的な雑談型音声情報案内システム.
言語処理学会年次大会発表論文集, C5-4, pp.761--764, 2014.

[28] 桑原暢弘, 秋田祐哉, 河原達也.
音声認識結果の有用性の自動判定に基づく 講義のリアルタイム字幕付与システム.
日本音響学会研究発表会講演論文集, 2-4-5, 春季 2014.

[29] Sheng Li, Yuya Akita, and Tatsuya Kawahara.
Data selection assisted by caption to improve acoustic modeling for lecture transcription.
日本音響学会研究発表会講演論文集, 2-4-4, 春季 2014.

[30] 秋田祐哉, 河原達也.
音声認識を用いたオンライン自動字幕作成・編集システム.
日本音響学会研究発表会講演論文集, 2-8-4, 秋季 2013.

[31] 北村大地, 猿渡洋, 中村哲, 高橋祐, 近藤多伸, 亀岡弘和.
Optimal divergence diversity for superresolution-based nonnegative matrix factorization.
日本音響学会研究発表会講演論文集, 3-2-9, 春季 2014.

[32] 室田勇騎, 北村大地, 中井駿介, 猿渡洋, 中村哲, 近藤多伸.
統計モデルパラメータ推定を用いたポストフィルタに基づく非負値行列因子分解の実験的評価.
日本音響学会研究発表会講演論文集, 2-1-5, 春季 2014.

[33] 吉江孝太郎, 猿渡洋, 中村哲.
両耳補聴システムにおける HRTF を利用した画像トラッキング併用型マルチモーダル・ブラインド音声抽出.
日本音響学会研究発表会講演論文集, 2-1-4, 春季 2014.

[34] 平野佑佳, 宮崎亮一, 猿渡洋, 中村哲.
ミュージカルノイズフリー音声抽出における音声カートシス比に基づく反復回数の制御.
日本音響学会研究発表会講演論文集, 2-1-3, 春季 2014.

[35] 宮内智, 北村大地, 猿渡洋, 中村哲.
音源到来方向分布とアクティベーション共有型非負値行列因子分解を用いた音像深度推定.
日本音響学会研究発表会講演論文集, 1-1-5, 春季 2014.

[36] 宮内智, 北村大地, 猿渡洋, 中村哲.
方位クラスタリングと非負値行列因子分解を用いた音像深度自動推定.
日本音響学会研究発表会講演論文集, 2-1-19, 秋季 2013.

[37] 平野佑佳, 宮崎亮一, 猿渡洋, 中村哲, 高谷智哉.
ブラインド音声抽出システムにおける音声認識率予測のための音声カートシス推定法.
日本音響学会研究発表会講演論文集, 2-1-6, 秋季 2013.

[38] 宮崎亮一, 猿渡洋, 中村哲, 近藤多伸.
様々なミュージカルノイズフリー音声強調法における音質評価.
日本音響学会研究発表会講演論文集, 2-1-2, 秋季 2013.

[39] 中井駿介, 宮崎亮一, 猿渡洋, 中村哲, 近藤多伸.
バイアス付き MMSE 短時間振幅スペクトル推定法の理論解析およびミュージカルノイズフリー雑音抑圧への拡張.
日本音響学会研究発表会講演論文集, 2-1-1, 秋季 2013.

[40] 北村大地, 猿渡洋, 中村哲, 近藤多伸, 高橋祐.
Divergence optimization based on trade-off between separation and extrapolation abilities in superresolution-based nonnegative matrix factorization.
日本音響学会研究発表会講演論文集, 1-1-6, 秋季 2013.

[41] 保澤圭亮, 吉本廣雅, 近藤一晃, 小泉敬寛, 中村裕一.
人間の視覚・運動特性を考慮した指差し支援インタフェース.
HCG シンポジウム, 2013.

[42] 小泉敬寛, 小幡佳奈子, 渡辺靖彦, 近藤一晃, 中村裕一.
映像対話型行動支援における作業者と支援者の態度の分析.
HCG シンポジウム, 2013.

[43] 松井研太, 近藤一晃, 小泉敬寛, 中村裕一.
個人視点映像を一覧するための広視野貼り合わせ画像群の自動生成 ～ 貼り合わせの良さに基づいた画像の選択とグループ化 ～.
HCG シンポジウム, 2013

[44] 保澤圭亮, 吉本廣雅, 近藤一晃, 小泉敬寛, 中村裕一.
動作の正確さと計測の精度に基づいた指差しインタフェース -確率密度によるポインティング表示-.
電子情報通信学会研究会報告, MVE2013-26, 2013.

[45] 松井研太, 近藤一晃, 小泉敬寛, 中村裕一.
個人視点映像からの広視野画像の自動生成 -輝度値の確率分布に基づいた貼り合わせに適した画像群の選択.
電子情報通信学会研究会報告, MVE2013-21, 2013.

[46] 近藤一晃, 森幹彦, 小泉敬寛, 中村裕一, 喜多一.
グループ学習における個人視点映像を用いた注目行動の自動認識に関する基礎調査.

情報処理学会情報教育シンポジウム, 2013.
[47] 松井研太, 近藤一晃, 小泉敬寛, 中村裕一.
輝度値の分布と情報量を用いた画像貼り合わせの評価.
電子情報通信学会研究会報告, PRMU2013-32, 2013.
[48] 村脇有吾.
フレーズベースTF-IDF: 名詞句解析の応用.
情報処理学会研究報告, NL-214-11, 2013.
[49] 河原達也.
議会の会議録作成のための音声認識－衆議院のシステムの概要－.
情報処理学会研究報告, SLP-93-5, 2012.
[50] Mijit Ablimit, Tatsuya Kawahara, and Askar Hamdulla.
Comparison of discriminative models for lexicon optimization for ASR of agglutinative language.
情報処理学会研究報告, SLP-92-13, 2012.
[51] Randy Gomez and Tatsuya Kawahara.
Wavelet packet decomposition approach to reverberant speech recognition.
情報処理学会研究報告, SLP-92-11, 2012.
[52] 吉野幸一郎, 森信介, 河原達也.
述語項構造を介した文の変換と選択に基づく音声対話用言語モデルの構築.
情報処理学会研究報告, SLP-91-3, NL-206-3, 2012.
[53] 秋田祐哉, 河原達也.
オープンコースウェアの講演を対象とした音声認識に基づく字幕付与.
日本音響学会研究発表会講演論文集, 2-9-9, 春季 2013.
[54] 秋田祐哉, 渡邉真人, 河原達也.
講演の音声認識と整形に基づく自動字幕付与.
日本音響学会研究発表会講演論文集, 1-1-18, 秋季 2012.
[55] 西村一馬, 川波弘道, 猿渡洋, 鹿野清宏.
音声情報案内システムのための統計的機械翻訳を利用した質問応答.
情報処理学会研究報告, SLP-91-14, NL-206-14, 2012.
[56] 糸井三由希, 宮崎亮一, 戸田智基, 猿渡洋, 鹿野清宏.
ユーザ動作に伴う雑音を含む非可聴つぶやき音声におけるブラインド音声抽出.
電子情報通信学会技術研究報告, EA2012-40, 2012.
[57] Suzumi Kanehara, Ryoichi Miyazaki, Hiroshi Saruwatari, Kiyohiro Shikano, and Kazunobu Kondo.
Mathematical Metric of Musical Noise for Various Nonlinear Speech Enhancement Algorithms.
IEICE Technical Report, EA2012-44, 2012.

[58] 真嶋温佳, 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 を特徴量とした無効入力の棄却.
情報処理学会研究報告, SLP-92-7, pp.1-6, 2012.

[59] 真嶋温佳, 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.
音声情報システムにおける最大エントロピー法を用いた無効入力棄却の評価.
日本音響学会講演論文集, 3-1-8, 2012.

[60] 金原涼美, 猿渡洋, 宮崎亮一, 鹿野清宏, 近藤多伸.
判定帰還型推定法に基づく音声強調法におけるミュージカルノイズ発生量の数理解析.
日本音響学会講演論文集, 2-9-3, 2012.

[61] 宮崎亮一, 猿渡洋, 鹿野清宏, 近藤多伸.
チャネル選択型 ICA に基づく雑音推定を用いたミュージカルノイズフリーブラインド音声強調.
日本音響学会講演論文集, 3-9-9, 2012.

[62] 糸井三由希, 宮崎亮一, 戸田智基, 猿渡洋, 鹿野清宏.
異種センサを用いて収録された非可聴つぶやき音声におけるブラインド音声抽出.
日本音響学会講演論文集, 3-9-10, 2012.

[63] 大沼侑司, 猿渡洋, 鹿野清宏.
マルチモーダル環境下における拡散性雑音を含む複数話者分離の高精度化.
日本音響学会講演論文集, 3-9-11, 2012.

[64] 大沼侑司, 猿渡洋, 鹿野清宏.
ポスター会議発表のマルチモーダルアーカイブを目的とした音源分離と評価.
信号処理シンポジウム, pp.412-417, 2012.

[65] 金原涼美, 猿渡洋, 宮崎亮一, 鹿野清宏, 近藤多伸.
様々な非線形音声強調法における近似モデルを用いたミュージカルノイズ発生量解析.
信号処理シンポジウム, pp.430-435, 2012.

[66] 宮崎亮一, 猿渡洋, 鹿野清宏, 近藤多伸.
様々な動的雑音推定器に基づくミュージカルノイズフリー雑音抑圧処理の評価.
信号処理シンポジウム, pp.436-441, 2012.

[67] 糸井三由希, 宮崎亮一, 戸田智基, 猿渡洋, 鹿野清宏.
ユーザ動作を含む非可聴つぶやき音声における多チャネル異種センサ統合に基づくブラインド音声抽出.
電子情報通信学会技術研究報告, EA2012-119, 2013.

[68] 真嶋温佳, 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 を用いた無効入力棄却モデルの可搬性の評価.
日本音響学会講演論文集, 3-9-5, 2013.

[69] 糸井三由希, 宮崎亮一, 戸田智基, 猿渡洋, 鹿野清宏.

ユーザ動作を含む非可聴つぶやき音声における多チャネル異種センサ統合に基づくブラインド音声抽出の評価.
日本音響学会講演論文集, 2-10-2, 2013
[70] 中井駿介, 大沼侑司, 宮崎亮一, 猿渡洋, 鹿野清宏.
ポスタ会議発表の音声アーカイブを目的とした音源分離の高音質化.
日本音響学会講演論文集, 2-10-3, 2013.
[71] 金原涼美, 猿渡洋, 宮崎亮一, 鹿野清宏, 近藤多伸.
様々な音声強調法におけるミュージカルノイズ発生量の数理解析の妥当性に関する検討.
日本音響学会講演論文集, 3-10-15, 2013.
[72] 高悠史, 吉本廣雅, 近藤一晃, 中村 裕一.
対話状況の可視化のためのヒューマン・コンピュータ協調モデル.
情報処理学会研究報告 HCI-148- 6, 2012.
[73] 近藤一晃, 松井研太, 中村裕一.
カメラ装着者の行動と閲覧時の注目対象に基づいた個人視点映像の加工.
画像の認識・理解シンポジウム論文集(MIRU), 2012.
[74] 吉本廣雅, 中野克己, 近藤一晃, 小泉敬寛, 中村裕一.
状況の認識とユーザの誘導を用いた協調的ジェスチャインタフェース.
画像の認識・理解シンポジウム論文集 (MIRU), 2012.
[75] 小泉敬寛, 中村裕一, 近藤一晃, 小幡佳奈子, 渡辺靖彦.
映像対話型行動記録におけるモダリティ間の関係と凝集性.
電子情報通信学会技術研究報告, HCS2012-33, 2012.
[76] 吉本廣雅・中村裕一.
ポスターセッションの分析のための不特定複数人物の頭部形状と姿勢のオンライン自動推定.
HCG シンポジウム, pp. 344-349, 2012.
[77] 高瀬恵三郎, 近藤一晃, 小泉敬寛, 中村裕一.
共同注視状況における複数人物頭部カメラの位置姿勢推定.
HCG シンポジウム, pp. 22-28, 2012.
[78] 朝倉僚, 宮坂淳介, 近藤一晃, 中村裕一, 秋田純一, 戸田真志, 櫻沢繁.
筋電位計測と kinect センサーによる三次元姿勢計測を用いたリハビリ支援システムの設計.
HCG シンポジウム, pp. 200-206, 2012.
[79] 村脇有吾, 黒橋禎夫.
名詞句の内部構造を考慮したキーワードのスコア付け.
言語処理学会年次大会, 2013 年 3 月.
[80] 岩立卓真, 高梨克也, 河原達也.
ポスター会話におけるパラ言語・非言語情報を用いた 話者交替及び次話者の予測.
人工知能学会研究会資料, SLUD-B103-10, 2012.

[81] 林宗一郎, 吉本廣雅, 平山高嗣, 河原達也.
マルチモーダルな認識に基づくポスター発表システム.
インタラクション, (インタラクティブ発表), pp.503--508, 2012.

[82] Welly Naptali and Tatsuya Kawahara.
Automatic transcription of ted talks.
音声ドキュメント処理ワークショップ, 2012.

[83] Mijit Ablimit, Tatsuya Kawahara, and Askar Hamdulla.
Evaluation of lexicon optimization based on discriminative learning.
情報処理学会研究報告, SLP-89-2, 2011.

[84] 渡邉真人, 秋田祐哉, 河原達也.
予稿の話し言葉変換に基づく言語モデルによる講演音声認識.
情報処理学会研究報告, SLP-89-1, 2011.

[85] 吉野幸一郎, 森信介, 河原達也.
述語項の類似度に基づいてニュース記事の案内を行う音声対話システム.
人工知能学会研究会資料, SLUD-B102-08, 2011.

[86] 平山高嗣, 角康之, 河原達也, 松山隆司.
情報コンシェルジェ:mind probing に基づくマルチモーダル インタラクションシステム.
電子情報通信学会技術研究報告, HCS2011-37, 2011.

[87] 土屋貴則, 高梨克也, 河原達也.
ポスター発表における質問者と質問の種類の推定のための マルチモーダルな聞き手行動分析.
人工知能学会研究会資料, SLUD-B101-14, 2011.

[88] Randy Gomez and Tatsuya Kawahara.
Robust speech recognition in noisy and reverberant environments using wavelet-based Wiener filtering.
情報処理学会研究報告, SLP-87-14, 2011.

[89] 吉野幸一郎, 森信介, 河原達也.
述語項の類似度に基づく情報推薦を行う音声対話システム.
情報処理学会研究報告, SLP-87-11, 2011.

[90] Cheongjae Lee, Tatsuya Kawahara, and Alexander Rudnicky.
Collecting speech data using Amazon's Mechanical Turk for evaluating voice search system.
情報処理学会研究報告, SLP-87-9, 2011.

[91] Mijit Ablimit, Tatsuya Kawahara, and Askar Hamdulla.
Lexicon optimization for automatic speech recognition based on discriminative learning.
情報処理学会研究報告, SLP-87-5, 2011.

[92] 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.
情報処理学会研究報告, SLP-87-4, 2011.

[93] 山口洋平, 森信介, 河原達也.
仮名漢字変換ログを用いた講義音声認識のための言語モデル適応.
言語処理学会年次大会発表論文集, C5-4, pp.1276--1279, 2012.

[94] 吉野幸一郎, 森信介, 河原達也.
述語項構造を用いた文変換とフィルタリングに基づく音声対話用言語モデル.
言語処理学会年次大会発表論文集, D3-2, pp.635--638, 2012.

[95] 渡邉真人, 秋田祐哉, 河原達也.
予稿の話し言葉変換に基づく言語モデルによる講演音声認識.
日本音響学会研究発表会講演論文集, 3-7-5, 春季 2012.

[96] 秋田祐哉, 河原達也.
講演における複数アノテーションに基づく句読点の自動挿入.
日本音響学会研究発表会講演論文集, 3-10-4, 秋季 2011.

[97] Ryo Wakisaka, Hiroshi Saruwatari, Kiyohiro Shikano, Tomoya Takatani.
Unsupervised parameter identification of MMSE STSA estimator.
IEICE Technical Report, SP2011-8, 2011.

[98] Ryoichi Miyazaki, Hiroshi Saruwatari, Kiyohiro Shikano.
Mathematical metric of speech distortion in various types of BSSA.
IEICE Technical Report, SP2011-9, 2011.

[99] 脇坂龍, 猿渡洋, 鹿野清宏, 高谷智哉.
一般化ガウス分布仮説とキュムラントの加法性を利用した雑音中からの音声カートシス逆推定.
信号処理シンポジウム, pp.320-325, 2011.

[100] 宮崎亮一, 猿渡洋, 鹿野清宏, 近藤多伸.
ミュージカルノイズフリー雑音抑圧の一般化理論とその信号抽出への応用.
信号処理シンポジウム, pp.368-373, 2011.

[101] 大沼侑司, 鎌土記良, 宮崎亮一, 猿渡洋, 鹿野清宏.
Kinect におけるリアルタイム・ブラインド空間サブトラクションアレーの実装と評価.
人工知能学会 AI チャレンジ研究会, B102-8, 2011.

[102] 岡本広大, 宮崎亮一, 猿渡洋, 鹿野清宏.
ポスタ会議発表音声アーカイブ構築を目的としたブラインド音声抽出の評価.
電子情報通信学会技術研究報告, EA2011-107, 2012.

[103] Ryoichi Miyazaki, Hiroshi Saruwatari, Kiyohiro Shikano, Kazunobu Kondo.
Evaluation of musical-noise-free noise reduction under real acoustic environments.
IEICE Technical Report, EA2011-108, 2012.

[104] Ryoichi Miyazaki, Hiroshi Saruwatari, Kiyohiro Shikano, Kazunobu Kondo.
Iterative blind spatial subtraction array for musical-noise-free speech enhancement in diffuse noise.
IEICE Technical Report, EA2011-125, 2012.

[105] 宮崎亮一, 猿渡洋, 井上貴之, 鹿野清宏, 近藤多伸.
ミュージカルノイズフリー雑音抑圧理論とその評価.
日本音響学会講演論文集, 1-4-2, 2011.

[106] 岡本広大, 宮崎亮一, 猿渡洋, 鹿野清宏.
ポスタ会議発表の音声アーカイブ構築を目的としたブラインド音声抽出.
日本音響学会講演論文集, 2-6-11, 2011.

[107] 脇坂龍, 猿渡洋, 鹿野清宏, 高谷智哉.
定位保持型 MMSE-STSA 推定に基づく両耳補聴システムの評価," 日本音響学会講演論文集, 3-6-7, 2011.

[108] 大沼侑司, 鎌土記良, 猿渡洋, 鹿野清宏.
Kinect を用いた話者位置トラッキングの併用による雑音抑圧処理の高精度化.
日本音響学会講演論文集, 1-1-16, 2012.

[109] 岡本広大, 宮崎亮一, 猿渡洋, 鹿野清宏.
ポスタ会議発表の音声アーカイブ構築を目的としたブラインド音声抽出と発話区間推定.
日本音響学会講演論文集, 1-1-17, 2012.

[110] 脇坂龍, 猿渡洋, 鹿野清宏, Frederic Mustiere, Martin Bouchard.
多チャネル MMSE-STSA 推定法を用いた定位保持型両耳補聴システムの評価.
日本音響学会講演論文集, 3-1-21, 2012.

[111] 宮崎亮一, 猿渡洋, 鹿野清宏, 近藤多伸.
ミュージカルノイズフリー雑音抑圧における音声歪み量の性能評価.
日本音響学会講演論文集, 3-1-22, 2012.

[112] 西村一馬, 川波弘道, 猿渡洋, 鹿野清宏.
統計的機械翻訳の手法を用いた音声情報案内システムのための応答文生成手法の検討.
情報処理学会研究報告, SLP-87-12, 2011.

[113] 西村一馬, 川波弘道, 猿渡洋, 鹿野清宏.
音声認識結果を用いた統計的機械翻訳による音声情報案内システム応答文の分析.
日本音響学会講演論文集, 3-7-13, 2012.

[114] 黒田真央, 延原章平, 松山隆司.
対称性制約を用いた多視点映像からの3次元顔形状復元と視線推定.
画像の認識・理解シンポジウム(MIRU) ,2011.

[115] 高橋康輔, 延原章平, 松山隆司.
参照物体の鏡像を用いた線形外部キャリブレーション法.
情報処理学会研究会資料, CVIM-180-25, 2012.

[116] 中野克己, 吉本廣雅, 近藤一晃, 小泉敬寛, 中村裕一.
ジェスチャインタフェースのユーザビリティ向上に向けたフィードバック構成.
電子情報通信学会技術研究報告 PRMU2011-182, 2012.
[117] 高悠史, 吉本廣雅, 近藤一晃, 中村裕一.
遠隔会議の実時間支援に向けた対話状況の可視化 ～対話の結束性に基づく表現の有効性～.
電子情報通信学会技術研究報告 MVE2011-2, 2011.
[118] 安光州, 近藤一晃, 小泉敬寛, 中村裕一.
個人視点映像を用いた対話シーンの検出・認識に関する検討.
電子情報通信学会技術研究報告 MVE2011-10, 2011.
[119] 柴田知秀, 村脇有吾, 黒橋禎夫, 河原大輔.
実テキスト解析をささえる語彙知識の自動獲得.
言語処理学会年次大会発表論文集, 2012.
[120] 村脇有吾, 岸本侑也, 黒橋禎夫.
ベイズ学習によるカタカナ複合語の分割.
言語処理学会年次大会発表論文集, 2012.
[121] Randy Gomez and Tatsuya Kawahara.
Robust speech recognition using optimized wavelet denoising with noise profiles.
情報処理学会研究報告, SLP-85-12, 2011.
[122] 秋田祐哉, 三村正人, Graham Neubig, 河原達也.
国会音声認識システムの音響・言語モデルの半自動更新.
情報処理学会研究報告, SLP-84-3, 2010.
[123] Randy Gomez and Tatsuya Kawahara.
Robust speech recognition using optimized wavelet filtering in reverberant conditions.
人工知能学会研究会資料, Challenge-B002-4, 2010.
[124] 吉野幸一郎, 河原達也.
Web からの情報抽出を用いた対話システムの評価.
人工知能学会研究会資料, SLUD-B002-04, 2010.
[125] 吉野幸一郎, 河原達也.
Web からの情報抽出を用いた音声対話システム.
情報処理学会研究報告, SLP-82-20, 2010.
[126] 須見康平, 河原達也.
音声会話コンテンツにおける聴衆の反応に基づいたホットスポットの抽出.
情報処理学会研究報告, SLP-82-8, 2010.
[127] 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.

情報処理学会研究報告, SLP-82-6, 2010.
[128] Randy Gomez and Tatsuya Kawahara.
Robust speech recognition using optimized wavelet-based dereverbaration.
情報処理学会研究報告, SLP-82-5, 2010.
[129] 河原達也, 秋田祐哉, 三村正人, 政瀧浩和, 高橋敏.
衆議院会議録作成における音声認識システム － 全体の構成と評価 －.
日本音響学会研究発表会講演論文集, 3-5-5, 春季 2011.
[130] 秋田祐哉, 河原達也, 政瀧浩和.
衆議院会議録作成における音声認識システム － 言語モデル －.
日本音響学会研究発表会講演論文集, 3-5-6, 春季 2011.
[131] 三村正人, 秋田祐哉, 河原達也.
衆議院会議録作成における音声認識システム － 音響モデル －.
日本音響学会研究発表会講演論文集, 3-5-7, 春季 2011.
[132] 吉野幸一郎, 森信介, 河原達也.
情報抽出と述語項の類似度を利用した音声対話システム.
言語処理学会年次大会発表論文集, D1-6, pp.107--110, 2011.
[133] 秋田祐哉, 河原達也.
講演における読点の個人的傾向のモデル化と自動挿入.
日本音響学会研究発表会講演論文集, 1-9-11, 秋季 2010.
[134] 久保慶伍, 川波弘道, 猿渡洋, 鹿野清宏.
未知語認識のための仮名・漢字単位の構築手法と性能評価.
情報処理学会研究報告, SLP-82-15, 2010.
[135] 久保慶伍, 川波弘道, 猿渡洋, 鹿野清宏,
多対多最小パターンアライメントアルゴリズムの提案と自動読み付与による評価.
情報処理学会研究報告, SLP-85-16, 2011.
[136] Takayuki Inoue, Hiroshi Saruwatari, Kiyohiro Shikano, and Kazunobu Kondo.
Mathematical metric of musical noise in Wiener filtering,.
信学技報, SP2010-105, 2011.
[137] 宮崎亮一, 井上貴之, 平田将久, 猿渡洋, 鹿野清宏, 高谷智哉.
非線形処理におけるミュージカルノイズ発生量と音声認識率の関係.
信学技報, SP2010-106, 2011.
[138] 宮崎亮一, 井上貴之, 平田将久, 猿渡洋, 鹿野清宏, 高谷智哉.
ブラインド雑音推定に基づく非線形雑音抑圧処理におけるミュージカルノイズ発生量と音声認識率の関係.
日本音響学会研究発表会講演論文集, 1-9-14, 春季 2011.
[139] 脇坂龍, 井上貴之, 猿渡洋, 鹿野清宏, 高谷智哉.
キュムラントの加法性を利用した雑音中からの音声カートシス逆推定.

日本音響学会研究発表会講演論文集, 2-9-6, 春季 2011.
[140] 井上貴之, 猿渡洋, 鹿野清宏, 近藤多伸.
非線形雑音抑圧手法におけるミュージカルノイズ発生量の数理解析.
日本音響学会研究発表会演論文集, 2-9-2, 春季 2011.
[141] 久保慶伍, 川波弘道, 猿渡洋, 鹿野清宏.
未知語の読み付与のための多対多最小パターンアライメント.
日本音響学会研究発表会講演論文集, 3-5-15, 春季 2011.
[142] 村脇有吾, 黒橋禎夫
混成型別サンプリングを用いた名詞句分割.
言語処理学会年次大会発表論文集, 2011.
[143] 勝木健太, 笹野遼平, 河原大輔, 黒橋禎夫.
Web 上の多彩な言語表現バリエーションに対応した頑健な形態素解析.
言語処理学会年次大会発表論文集, 2011.
[144] 荒井翔真, 柴田知秀, 黒橋禎夫.
発表スライドの言語的・構造的解釈に基づく発話生成.
言語処理学会年次大会発表論文集, 2011.
[145] 近藤一晃, 西谷英之, 中村裕.
協調的物体認識のためのインタラクション設計.
画像の認識・理解シンポジウム論文集(MIRU), 2010.
[146] 中野克己, 吉本廣雅, 近藤一晃, 小泉敬寛, 中村裕一.
ジェスチャインタフェースにおける画像認識とフィードバックの構成論.
HCG シンポジウム, C4-2, 2010.
[147] 高梨克也, 常志強, 河原達也.
聞き手の興味・関心を示すあいづちの生起する会話文脈の分析.
人工知能学会研究会資料, SLUD-A903-05, 2010.
[148] Graham Neubig, 秋田祐哉, 森信介, 河原達也.
文脈を考慮した確率的モデルによる話し言葉の整形.
情報処理学会研究報告, SLP-79-17, 2009.
[149] 須見康平, 河原達也, 緒方淳, 後藤真孝.
Podspotter: 音リアクションイベント検出に基づくポッドキャストブラウザ.
WISS (インタラクティブシステムとソフトウェアに関するワークショップ), pp.171--172, 2009.
[150] Randy Gomez and Tatsuya Kawahara.
Speech enhancement optimization based on acoustic model likelihood for noisy and reverberant environment.
人工知能学会研究会資料, Challenge-A902-9, 2009.
[151] 須見康平, 河原達也.
ポスター会話中の音リアクションイベントに基づくホットスポットの抽出.

情報処理学会全国大会講演論文集, 第 5 巻, pp.127--128, 2010.
[152] 吉野幸一郎, 河原達也.
Web からの情報抽出に基づく雑談的な対話の生成.
言語処理学会年次大会発表論文集, C2-2, pp.214--217, 2010.
[153] Graham Neubig, 秋田祐哉, 森信介, 河原達也.
統計的機械翻訳の枠組みを用いた話し言葉の整形.
言語処理学会年次大会発表論文集, C2-4, pp.222--225, 2010.
[154] 秋田祐哉, 河原達也.
講演の書き起こしに対する読点の自動挿入.
日本音響学会研究発表会講演論文集, 2-6-11, 春季 2010.
[155] 藤田洋子, 竹内翔大, 川波弘道, 松井知子, 猿渡洋, 鹿野清宏.
単語の頻度と音響の特徴を利用した SVM による無効入力の棄却.
情報処理学会研究報告, SLP-80-3, 2010.
[156] 藤田洋子, 竹内翔大, 川波弘道, 松井知子, 猿渡洋, 鹿野清宏.
Bag-of-Words を用いた SVM による無効発話の棄却.
日本音響学会研究発表会講演論文集, 3-6-16, 春季 2010.
[157] 岡本亮維, 高橋祐, 猿渡洋, 鹿野清宏.
独立成分分析を用いた MMSE STSA 推定法における目的音統計モデルのブラインド適応.
日本音響学会研究発表会講演論文集, 2-5-11, 春季 2010.
[158] 井上貴之, 高橋祐, 石川陽平, 猿渡洋, 鹿野清宏, 近藤多伸.
一般化スペクトル減算法におけるミュージカルノイズ発生量の数理解析.
日本音響学会研究発表会講演論文集, 3-5-4, 春季 2010.
[159] 高橋祐, 猿渡洋, 鹿野清宏, 近藤多伸.
スペクトル減算とアレー信号処理の 統合手法におけるミュージカルノイズ発生量の高次統計量に基づく数理解析の一般化.
日本音響学会研究発表会講演論文集, 3-5-5, 春季 2010.
[160] 石川陽平, 高橋祐, 猿渡洋, 鹿野清宏, 近藤多伸.
ミュージカルノイズ制御型アレー信号処理手法の実環境評価.
日本音響学会研究発表会講演論文集, 3-5-6, 春季 2010.
[161] 原島純, 黒橋禎夫.
PLSI を用いたウェブ検索結果の要約.
言語処理学会年次大会発表論文集, 2010.
[162] 村脇有吾, 黒橋禎夫.
テキストから自動獲得した名詞の分類.
言語処理学会年次大会発表論文集, 2010.
[163] 西谷英之, 近藤一晃, 中村裕一.

人間との協調による物体認識のためのインタラクション設計.
信学技報 PRMU2009-308, pp. 443-448, 2010.
[164] 高悠史, 近藤一晃, 小泉敬寛, 中村裕一.
遠隔会議の実時間支援のためのスナップショット取得と共有.
信学技報 MVE2009-62, pp.21-22, 2009.
[165] 西谷英之, 近藤一晃, 中村裕一.
人間とのインタラクションを用いた協調的物体認識.
信学技報 MVE2009-59, pp.7-12, 2009.

（国際）

[1] T.Kawahara, M.Uesato, K.Yoshino, and K.Takanashi.
Toward adaptive generation of backchannels for attentive listening agents.
Int'l Workshop Spoken Dialogue Systems (IWSDS), Busan, Korea, January, 2015.

[2] K.Yoshino and T.Kawahara.
News navigation system based on proactive dialogue strategy.
Int'l Workshop Spoken Dialogue Systems (IWSDS), Busan, Korea, January, 2015.

[3] Y.Wakabayashi, K.Inoue, H.Yoshimoto, and T.Kawahara.
Speaker diarization based on audio-visual integration for smart posterboard.
APSIPA ASC, Siem Riep, Cambodia, December, 2014.

[4] K.Yoshino and T.Kawahara.
Information navigation system based on POMDP that tracks user focus.
SIGdial Meeting Discourse & Dialogue, Philadelphia, USA, June, 2014.

[5] M.Mimura, S.Sakai, and T.Kawahara.
Exploring deep neural networks and deep autoencoders in reverberant speech recognition.
Workshop on Hands-free Speech Communication & Microphone Arrays (HSCMA), Nancy, France, May, 2014.

[6] F.Aprilyanti, H.Saruwatari, K.Shikano, S.Nakamura, T.Takatani.
Optimized joint noise suppression and dereverberation based on blind signal extraction for hands-free speech recognition system.
Workshop on Hands-free Speech Communication & Microphone Arrays (HSCMA), Nancy, France, May, 2014.

[7] M.Yoshimoto and Y.Nakamura.
Cooperative Gesture Recognition: Learning Characteristics of Classifiers and Navigating User to Ideal Situation.
IEEE Int'l Conf. Pattern Recognition Applications and Methods, Lisbon, Portugal, January, 2015.

[8] Y.Nakamura, T.Koizumi, K.Obata, K.Kondo, abd Y.Watanabe.

Behaviors and Communications in Working Support through First Person Vision Communication.
IEEE Int'l Conf. Ubiquitous Intelligence and Computing, Bali, Indonesia, December, 2014.

[9] T.Kawahara, S.Hayashi, and K.Takanashi.
Estimation of interest and comprehension level of audience through multi-modal behaviors in poster conversations.
INTERSPEECH, Lyon, France, August 2013.

[10] K.Yoshino, S.Mori, and T.Kawahara.
Incorporating semantic information to selection of web texts for language model of spoken dialogue system.
IEEE-ICASSP, Vancouver, Canada, May 2013.

[11] S.Nakai, R.Miyazaki, H.Saruwatari, and S.Nakamura.
Theoretical analysis of musical noise generation for blind speech extraction with generalized MMSE short-time spectral amplitude estimator.
Intelligent Signal Processing (ISP) Conf., London, UK, December 2013.

[12] H.Saruwatari and R.Miyazaki.
Information-geometric optimization for nonlinear noise reduction systems.
Int'l Sympo. Intelligent Signal Processing and Communication Systems (ISPACS), Naha, Japan, November 2013.

[13] R.Miyazaki, H.Saruwatari, S.Nakamura, K.Shikano, and K.Kondo, J.Blanchette, and M.Bouchard.
Toward musical-noise-free blind speech extraction: concept and its applications.
APSIPA ASC, Kaohsiung, Taiwan, October 2013.

[14] H.Saruwatari, S.Kanehara, R.Miyazaki, K.Shikano, K.Kondo.
Musical noise analysis for Bayesian minimum mean-square error speech amplitude estimators based on higher-order statistics.
INTERSPEECH, Lyon, France, August 2013.

[15] H.Yoshimoto and Y.Nakamura.
Cubistic Representation for Real-Time 3D Shape and Pose Estimation of Unknown Rigid Object.
ICCV Workshop, Sydney, Australia, December 2013.

[16] H.Yoshimoto and Y.Nakamura.
Free-Angle 3D Head Pose Tracking Based on Online Shape Acquisition.
ACPR, Naha, Japan, November 2013.

[17] T.Tung, R. Gomez, T. Kawahara, and T.Matsuyama.
Multi-party Human-Machine Interaction Using a Smart Multimodal Digital

Signage.
HCI, Las Vegas, USA, July 2013.

[18] Y.Murawaki.
Global Model for Hierarchical Multi-Label Text Classification.
IJCNLP, Nagoya, Japan, October 2013.

[19] K.Yoshino, S.Mori, and T.Kawahara.
Language modeling for spoken dialogue system based on filtering using predicate-argument structures.
COLING, Mumbai, India, December 2012.

[20] C.Lee and T.Kawahara.
Hybrid vector space model for flexible voice search.
APSIPA ASC, Hollywood, USA, December 2012.

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

[22] T.Kawahara.
Transcription system using automatic speech recognition for the Japanese Parliament (Diet).
AAAI/IAAI, Toronto, Canada, July 2012.

[23] R.Miyazaki, H.Saruwatari, K.Shikano, and K.Kondo.
Musical-noise-free speech enhancement based on iterative Wiener filtering.
IEEE Int'l Sympo. Signal Processing & Information Technology (ISSPIT), Ho Chi Minh City, Vietnam, December, 2012.

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

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

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

[27] H.Saruwatari, R.Wakisaka, K.Shikano, F.Mustiere, L.Thibault, H.Najaf-Zadeh,

and M.Bouchard.
Sound-localization-preserved binaural MMSE STSA estimator with explicit and implicit binaural cues.
EUSIPCO, Bucharest, Romania, August 2012.

[28] T.Tung, R.Gomez, T.Kawahara, and T.Matsuyama.
Group dynamics and multimodal interaction modeling using a smart digital signage.
ECCV Workshop Video Event Categorization, Tagging and Retrieval, Florence, Italy, October 2012.

[29] Y.Murawaki and S.Kurohashi.
Semi-Supervised Noun Compound Analysis with Edge and Span Features.
COLING, Mumbai, India, December 2012.

[30] J.Harashima and S.Kurohashi.
Flexible Japanese sentence compression by relaxing unit constraints.
COLING, Mumbai, India, December 2012.

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

[32] T.Hirayama, Y.Sumi, T.Kawahara, and T.Matsuyama.
Info-concierge: Proactive multi-modal interaction using mind probing.
APSIPA ASC, Xian, China, October 2011.

[33] H.Saruwatari, N.Hirata, T.Hatta, R.Wakisaka, K.Shikano, T.Takatani.
Semi-blind speech extraction for robot using visual information and noise statistics.
IEEE ISSPIT, pp.238-243, Bilbao, Spain, December, 2011.

[34] R.Miyazaki, H.Saruwatari, K.Shikano.
Theoretical analysis of musical noise and speech distortion in structure-generalized parametric blind spatial subtraction array.
INTERSPEECH, pp.341-344, Florence, Italy, August 2011.

[35] R.Wakisaka, H.Saruwatari, K.Shikano, T.Takatani.
Blind speech prior estimaion for generalized minimum mean-square error short-time spectral amplitude estimator.
INTERSPEECH, pp.361-364, Florence, Italy, August 2011.

[36] Z.Yu, X.Zhou, Z.Yu, C.Becker, and Y.Nakamura.
Social Interaction Mining in Small Group Discussion Using a Smart Meeting System.

UIC, Banff, Canada, September 2011.

[37] Y.Murawaki and S.Kurohashi.
Non-parametric Bayesian Segmentation of Japanese Noun Phrases.
EMNLP, Edinburgh, UK, July 2011.

[38] Y.Akita, M.Mimura, G.Neubig, and T.Kawahara.
Semi-automated update of automatic transcription system for the Japanese national congress.
INTERSPEECH, Makuhari, September 2010.

[39] T.Tung and T.Matsuyama.
3D Video Performance Segmentation.
IEEE-ICIP, Hong Kong, September 2010.

[40] P.Huang, T.Tung, S. Nobuhara, A.Hilton, and T. Matsuyama.
Comparison of Skeleton and Non-Skeleton Shape Descriptors for 3D Video.
Int'l Sympo. 3D Data Processing, Visualization & Transmission (3DPVT), Paris, France, May 2010.

[41] K.Kondo, H.Nishitani, and Y.Nakamura
Human-Computer Collaborative Object Recognition for Intelligent Support.
Pacific-Rim Conference on Multimedia (PCM), Shanghai, China, September 2010.

③ ポスター発表 （国内会議 30 件、国際会議 40 件）

（国内）

[1] 若林佑幸, 中山雅人, 西浦敬信, 山下洋一, 井上昂治, 吉本廣雅, 河原達也.
拡散性雑音環境下における多人数会話のマルチモーダル話者区間検出.
日本音響学会研究発表会講演論文集, 1-Q-24, 春季 2015.

[2] Sheng Li, Yuya Akita, and Tatsuya Kawahara.
Incorporating divergences from hypotheses of multiple ASR systems to improve unsupervised acoustic model training.
日本音響学会研究発表会講演論文集, 1-P-23, 春季 2015.

[3] 三村正人, 坂井信輔, 河原達也.
音素クラス情報を用いたディープオートエンコーダによる残響下音声認識.
日本音響学会研究発表会講演論文集, 1-P-22, 春季 2015.

[4] 三村正人, 坂井信輔, 河原達也.
ディープオートエンコーダと DNN-HMM を用いた残響下音声認識.
日本音響学会研究発表会講演論文集, 1-R-5, 秋季 2014.

[5] Sheng Li, Yuya Akita, and Tatsuya Kawahara.
Unsupervised training of deep neural network acoustic models for lecture transcription.

日本音響学会研究発表会講演論文集, 1-R-4, 秋季 2014.
[6] 中井駿介, 宮崎亮一, 猿渡洋, 中村哲, 井上昂治, 若林佑幸, 河原達也.
スマートポスターボードにおける実環境を想定した複数話者分離.
日本音響学会研究発表会講演論文集, 2-Q4-8, 春季 2014.
[7] 若林佑幸, 井上昂治, 河原達也, 中井駿介, 宮崎亮一, 猿渡洋.
スマートポスターボードにおける音響情報と画像情報の統合による 話者区間検出.
日本音響学会研究発表会講演論文集, 2-Q4-7, 春季 2014.
[8] 三村正人, 河原達也.
講演音声認識における DNN-HMM の教師なし話者適応.
日本音響学会研究発表会講演論文集, 2-Q4-22, 春季 2014.
[9] Sheng Li, Masato Mimura, and Tatsuya Kawahara.
Automatic transcription of Chinese spoken lectures.
日本音響学会研究発表会講演論文集, 2-P-31, 秋季 2013.
[10] 三村正人, 河原達也.
CSJ を用いた日本語講演音声認識用 DNN-HMM の構築.
日本音響学会研究発表会講演論文集, 1-P-42b, 秋季 2013.
[11] 大沼侑司, 中井駿介, 宮崎亮一, 猿渡洋, 鹿野清宏.
ポスタ会議発表マルチモーダルアーカイブ構築のための音声区間検出.
日本音響学会講演論文集, 1-P-35, 2013.
[12] 宮崎亮一, 猿渡洋, 鹿野清宏, 近藤 多伸.
非線形信号処理後の信号に対する雑音推定精度の検討.
日本音響学会講演論文集, 1-P-45, 2013.
[13] 近藤一晃, 松井研太, 中村裕一.
カメラ装着者の行動と閲覧時の注目対象に基づいた個人視点映像の加工.
画像の認識・理解シンポジウム論文集(MIRU2012), 2012.
[14] 吉本廣雅, 中野克己, 近藤一晃, 小泉敬寛, 中村裕一.
状況の認識とユーザの誘導を用いた協調的ジェスチャインタフェース.
第 15 回画像の認識・理解シンポジウム論文集(MIRU2012), 2012.
[15] Welly Naptali and Tatsuya Kawahara.
Automatic speech recognition for ted talks.
日本音響学会研究発表会講演論文集, 3-P-4, 春季 2012.
[16] Randy Gomez and Tatsuya Kawahara.
Wavelet packet decomposition-based dereverberation for robust ASR.
日本音響学会研究発表会講演論文集, 1-P-16, 春季 2012.
[17] 三村正人, 河原達也.
大学講義の音声認識のための音響・言語モデル適応に関する検討.
日本音響学会研究発表会講演論文集, 3-P-6, 秋季 2011.

[18] Randy Gomez and Tatsuya Kawahara.
Robust speech recognition in noisy and reverberant conditions using Wiener filtering in the wavelet.
日本音響学会研究発表会講演論文集, 2-Q-21, 秋季 2011.

[19] 久保慶伍, 川波弘道, 猿渡洋, 鹿野清宏.
発音付与のための EM アルゴリズムを用いた多対多アライメントの評価.
日本音響学会講演論文集, 3-P-8, 2012.

[20] 西村一馬, 川波弘道, 猿渡洋, 鹿野清宏.
音声情報案内システムにおける統計的機械翻訳の手法を用いた応答文生成手法の検討.
日本音響学会講演論文集, 3-P-25, 2011.

[21] 平井良佑, 久保慶伍, 木佐木雄介, 川波弘道, 猿渡洋, 鹿野清宏.
遷都 1300 年祭会場における音声情報案内システムの運用と発話データの分析.
日本音響学会講演論文集, 3-P-24, 2011.

[22] 平井良佑, 竹内翔大, 川波弘道, 猿渡洋, 鹿野清宏.
音声認識結果による類似スコアを用いた質問応答データベース拡張コストの削減.
日本音響学会講演論文集, 3-P-27, 2012.

[23] Randy Gomez and Tatsuya Kawahara.
Wavelet optimization using noise profiles for noise-robust speech recognition.
日本音響学会研究発表会講演論文集, 2-P-17, 春季 2011.

[24] 三村正人, 河原達也.
BIC に基づく話者モデル選択による高速な話者正規化と話者適応.
日本音響学会研究発表会講演論文集, 1-Q-21, 秋季 2010.

[25] Randy Gomez and Tatsuya Kawahara.
Wavelet optimization for robust dereverberation in automatic speech recognition.
日本音響学会研究発表会講演論文集, 1-Q-8, 秋季 2010.

[26] 吉田仙, 高梨克也, 河原達也, 永田昌明. ポスター会話における指示表現の分析 —参照先との自動対応付けに向けて—. 言語処理学会年次大会発表論文集, PB1-3, pp.430--433, 2010.

[27] 三村正人, 河原達也. 会議音声認識における発話の区分化と話者正規化の高速化. 日本音響学会研究発表会講演論文集, 2-Q-11, 春季 2010.

[28] Randy Gomez and Tatsuya Kawahara. Wavelet filtering in ASR robust to noisy and reverberant environments. 日本音響学会研究発表会講演論文集, 1-Q-2, 春季 2010.

[29] Rafael Torres, Shota Takeuchi, Hiromichi Kawanami ,Tomoko Matsui , Hiroshi Saruwatari, Kiyohiro Shikano. PrefixSpan Boosting-based inquiry classification for a speech-oriented guidance system. 日本音響学会研究発表会講演論文集, 1-Q-31, 春季 2010.

[30] 村脇有吾, 黒橋禎夫. オンライン語彙獲得を用いたリアルタイムウェブの言語処理. 言語処理学会年次大会発表論文集, 2010.

（国際）

[1] M.Mimura and T.Kawahara.
Unsupervised speaker adaptation of DNN-HMM by selecting similar speakers for lecture transcription.
APSIPA ASC, Siem Riep, Cambodia, December, 2014.

[2] K.Inoue, Y.Wakabayashi, H.Yoshimoto, and T.Kawahara.
Speaker diarization using eye-gaze information in multi-party conversations.
INTERSPEECH, Singapore, September, 2014.

[3] S.Li, Y.Akita, and T.Kawahara.
Corpus and transcription system of Chinese Lecture Room.
Int'l Sympo. Chinese Spoken Language Processing (ISCSLP), Singapore, September, 2014.

[4] S.Nakai, H.Saruwatari, R.Miyazaki, S.Nakamura, K.Kondo.
Theoretical analysis of biased MMSE short-time spectral amplitude estimator and its extension to musical-noise-free speech enhancement.
Workshop on Hands-free Speech Communication & Microphone Arrays (HSCMA), Nancy, France, May, 2014.

[5] T. Tung and T. Matsuyama.
Timing-based Local Descriptor for Dynamic Surfaces.
IEEE-CVPR, Columbus, USA, June, 2014.

[6] K.Yoshino, S.Mori, and T.Kawahara.
Predicate argument structure analysis using partially annotated corpora.
IJCNLP, Nagoya, October 2013.

[7] T.Tung and T.Matsuyama.
Intrinsic Characterization of Dynamic Surfaces.
IEEE-CVPR, Portland, USA, June 2013.

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

[9] R.Gomez and T.Kawahara.
Dereverberation based on wavelet packet filtering for robust automatic speech recognition.
INTERSPEECH, Portland, USA, September 2012.

[10] T.Kawahara, T.Iwatate, and K.Takanashi.

Prediction of turn-taking by combining prosodic and eye-gaze information in poster conversations.
INTERSPEECH, Portland, USA, September 2012.

[11] 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?
Interdisciplinary Workshop on Feedback Behaviors in Dialog, Stevenson, USA, September 2012.

[12] M.Ablimit, T.Kawahara, and A.Hamdulla.
Discriminative approach to lexical entry selection for automatic speech recognition of agglutinative language.
IEEE-ICASSP, Kyoto, March 2012.

[13] K.Nishimura, H.Kawanami, H.Saruwatari, and K.Shikano.
Response generation based on statistical machine translation for speech-oriented guidance system.
APSIPA ASC, Hollywood, USA, December 2012.

[14] 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.
APSIPA ASC, Hollywood, USA, December 2012.

[15] S.Hara, H.Kawanami, H.Saruwatari, and K.Shikano.
Development of a toolkit handling multiple speech-oriented guidance agents for mobile applications.
Int'l Workshop Spoken Dialog Systems (IWSDS), Paris, France, November 2012.

[16] 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.
Int'l Workshop Spoken Dialog Systems (IWSDS), Paris, France, November, 2012.

[17] 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.
INTERSPEECH, Portland, USA, September 2012.

[18] R.Miyazaki, H.Saruwatari, K.Shikano, and K.Kondo.
Musical-noise-free blind speech extraction using ICA-based noise estimation with channel selection.
Int'l Workshop Acoustic Signal Enhancement (IWAENC), Aachen, Germany,

September 2012.

[19] 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.
Int'l Workshop Acoustic Signal Enhancement (IWAENC), Aachen, Germany, September 2012.

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

[21] R.Miyazaki, H.Saruwatari, T.Inoue, K.Shikano, K.Kondo.
Musical-noise-free speech enhancement: theory and evaluation.
IEEE-ICASSP, Kyoto, March 2012.

[22] T.Tung and T.Matsuyama.
Invariant Surface-Based Shape Descriptor for Dynamic Surface Encoding,
ACCV (LNCS), Daejeon, Korea, Novermber 2012.

[23] R.Gomez and T.Kawahara.
Optimized wavelet-based speech enhancement for speech recognition in noisy and reverberant conditions.
APSIPA ASC, Xian, China, October 2011.

[24] M.Mimura and T.Kawahara.
Fast speaker normalization and adaptation based on BIC for meeting speech recognition.
APSIPA ASC, Xian, China, October 2011.

[25] M.Ablimit, T.Kawahara, and A.Hamdulla.
Lexicon optimization for automatic speech recognition based on discriminative learning.
APSIPA ASC, Xian, China, October 2011.

[26] Y.Akita and T.Kawahara.
Automatic comma insertion of lecture transcripts based on multiple annotations.
INTERSPEECH, Florence, Italy, August 2011.

[27] R.Gomez and T.Kawahara.
Denoising using optimized wavelet filtering for automatic speech recognition.
INTERSPEECH, Florence, Italy, August 2011.

[28] K.Kubo, H.Kawanami, H.Saruwatari, K.Shikano.
Unconstrained many-to-many alignment for automatic pronunciation annotation.

APSIPA ASC, Xian, China, October 2011.

[29] Y.Fujita, S.Takeuchi, H.Kawanami, T.Matsui, H.Saruwatari, K.Shikano.
Out-of-task utterance detection based on bag-of-words using automatic speech recognition results.
APSIPA ASC, Xian, China, October 2011.

[30] K.Nishimura, H.Kawanami, H.Saruwatari, K.Shikano.
Investigation of statistical machine translation applied to answer generation for a speech-oriented guidance system.
APSIPA ASC, Xian, China, October 2011.

[31] T.Inoue, H.Saruwatari, K.Shikano, K.Kondo.
Theoretical analysis of musical noise in Wiener filtering family via higherorder statistics.
IEEE-ICASSP, Prague, Czech, May 2011.

[32] R.Gomez and T.Kawahara.
Optimizing wavelet parameters for dereverberation in automatic speech recognition.
APSIPA ASC, Singapore, December 2010.

[33] T.Kawahara, K.Sumi, Z.Q.Chang, and K.Takanashi.
Detection of hot spots in poster conversations based on reactive tokens of audience.
INTERSPEECH, Makuhari, September 2010.

[34] T.Kawahara, N.Katsumaru, Y.Akita, and S.Mori.
Classroom note-taking system for hearing impaired students using automatic speech recognition adapted to lectures.
INTERSPEECH, Makuhari, September 2010.

[35] R.Gomez and T.Kawahara.
An improved wavelet-based dereverberation for robust automatic speech recognition.
INTERSPEECH, Makuhari, September 2010.

[36] T.Kawahara, Z.Q.Chang, and K.Takanashi.
Analysis on prosodic features of Japanese reactive tokens in poster conversations.
Int'l Conf. Speech Prosody, Chicago, USA, May 2010.

[37] T.Inoue, H.Saruwatari, K.Shikano, and K.Kondo.
Theoretical analysis of musicial noise in Wiener filter via higher-order statistics.
APSIPA ASC, Singapore, December 2010.

[38] Y.Murawaki and S.Kurohashi.
Online Japanese Unknown Morpheme Detection using Orthographic Variation.
Int'l Conf. Language Resources & Evaluation (LREC), Malta, June 2010.

[39] Y.Murawaki and S.Kurohashi
Semantic Classification of Automatically Acquired Nouns using Lexico-Syntactic Clues.
COLING, Poster Volume, Beijing, China, August 2010.

[40] T.Tung and T.Matsuyama.
Dynamic Surface Matching by Geodesic Mapping for 3D Animation Transfer.
IEEE-CVPR, San Francisco, USA, June 2010.

（4） 知財出願

① 国内出願（0 件）

なし

② 海外出願（0 件）

なし

③ その他知的財産権

なし

（5） 受賞・報道等

① 受賞

[1] 猿渡洋.
*** 2015 年度 科学技術分野の文部科学大臣表彰 科学技術賞（研究部門）.**
音響メディアにおける統計的信号処理の先駆的研究.

[2] 井上昂治.
情報処理学会第 77 回全国大会 学生奨励賞.
スマートポスターボードにおける視線情報を用いた話者区間検出及び相槌の同定.

[3] 吉本廣雅, 中村裕一.
HCG シンポジウム 2014 インタラクティブ発表賞.
ポスター自動発表システムの実現に向けた視線分布に基づくポスター対話の話題推定.

[4] 河原達也.
*** 2012 年度 情報処理学会 論文賞.**
音声会話コンテンツにおける聴衆の反応に基づく音響イベントとホットスポットの検出.

[5] 猿渡洋.
*** 2012 年度 新技術開発財団 市村学術賞・功績賞.**
高次統計量追跡に基づくブラインド音声抽出およびその高品質化.

[6] 宮崎亮一, 猿渡洋, 鹿野清宏.

APSIPA ASC 2013 Best Paper Award.
Toward musical-noise-free blind speech extraction: concept and its applications.

[7] 政瀧浩和, 堀貴明, 小橋川哲, 河原達也.
2012 年度 前島密賞(研究開発).
音声認識議会録作成システムの研究開発と実用化.

[8] 河原達也.
*** 2012 年度 ドコモ・モバイル・サイエンス賞(先端技術部門).**
話し言葉の音声認識に関する研究開発.

[9] 河原達也.
2012 IAAI Deployed Application Award.
Transcription System using Automatic Speech Recognition for the Japanese Parliament (Diet).

[10] 河原達也, 秋田祐哉, 三村正人.
*** 2012 年度 科学技術分野の文部科学大臣表彰 科学技術賞(研究部門).**
国会審議の自動音声認識システムの研究

[11] 河原達也, 秋田祐哉, 三村正人, 堀貴明, 小橋川哲.
2011 年度 情報処理学会 喜安記念業績賞.
議会の会議録作成のための音声認識システムの実用化.

[12] 猿渡洋.
*** 2011 年度 ドコモ・モバイル・サイエンス賞(基礎科学部門).**
ブラインド音源分離に基づく自律的な音声情報通信インターフェイスの先駆的研究.

[13] 井上貴之, 猿渡洋, 高橋祐, 鹿野清宏, 近藤多伸.
2011 年度 電気通信普及財団テレコムシステム技術賞.
Theoretical analysis of musical noise in generalized spectral subtraction based on higher-order statistics. (IEEE Trans. Audio, Speech and Language Processing).

② マスコミ(新聞・TV等)報道

[1] 2013 年 12 月 16 日 読売新聞(科学面)
音声認識 学術情報メディアセンター 河原達也教授

[2] 2013 年 2 月 23 日 京都新聞(教育面)
講義や講演 瞬時に字幕化 京大メディアセンター 実用化へ精度向上

[3] 2012 年 4 月 19 日 読売新聞(夕刊)
音声認識 精度アップ

(以下は 2011 年 5 月 12 日にプレスリリースしたもの)

[4] 2011 年 5 月 13 日 日本経済新聞
京大開発の音声認識システム、衆院が議事録作成に導入

[5] 2011 年 5 月 13 日 産経新聞
難解な専門用語もお任せ 世界初の国会答弁認識システム、京大教授ら開発 衆院で運用開始

[6] 2011 年 5 月 13 日 朝日新聞(関西版)
衆議院の議事録作成に音声認識システム 京大開発の技術

[7] 2011 年 5 月 13 日 京都新聞
余分な言葉除去、国会答弁自動で文字化 京大開発

[8] 2011 年 5 月 13 日 東京新聞
国会答弁の音声認識システム開発 京都大、正式運用を開始

[9] 2011 年 5 月 13 日 日刊工業新聞
京大の自動音声認識技術 国会で活躍 手書き速記に代わり衆院の会議録を作成

[10] 2011 年 5 月 13 日 ウォールストリートジャーナル日本版(電子版)
音声認識で議事録作成 衆院導入、世界初

[11] 2011 年 5 月 15 日 毎日新聞(大阪版)
国会議事録:機械まかせ 衆院、音声認識導入 「えー」自動で削除／新語も学習

[12] 2011 年 5 月 30 日 読売新聞(京都山城面)
音声認識で衆院議事録 京大技術採用 PC に文書表示

[13] 2011 年 8 月 16 日 毎日新聞(東京版夕刊)
国会「速記」に音声変換 効率化、衆院で世界初導入

2011 年度に音声認識技術を導入した衆議院の会議録作成システムが正式運用となったのを機に報道発表を行ったところ、上記の通り新聞主要各紙の他に、NHK 全国ニュースなどで取り上げられた。その後も、雑誌やテレビ番組などの取材が相次いだ。

③ その他

サイエンスアゴラ 2014(http://www.jst.go.jp/csc/scienceagora/)で展示を行った。

(6) 成果展開事例

① 実用化に向けての展開

- **議会審議のための音声認識技術は衆議院で導入され、2011 年度から全面的に利用されている。**
- 同上の技術は、NTT へのライセンス供与を通して、地方議会向けのパッケージにもなり、既にいくつかの議会で導入されている。
- この他に、大手速記会社などにライセンス供与を行っている。

② 社会還元的な展開活動

- 音声認識技術を用いて国会審議映像に字幕を付与する技術を、政策研究大学院大学の比

較議会情報プロジェクトへ提供している。

- **音声認識を用いて講演映像コンテンツに字幕を付与する技術を、京都大学 OCW (OpenCourseWare) の講演動画の字幕付与に応用している。一部の講演に対しては既に公式の字幕として配信されている。**
- **さらに、放送大学の講義の字幕付与への展開を進めている。**
- 最終的には、上記で述べた講演や会議の音声を自動で書き起こし、字幕付与するシステムをクラウドサーバ上で提供する予定である。
  (http://caption.ar.media.kyoto-u.ac.jp)
- 音声認識を用いた情報保障技術は、2015 年度に発足する情報処理学会のアクセシビリティ研究グループの活動へも貢献したいと考えている。

§5 研究期間中の活動

5.1 主なワークショップ、シンポジウム、アウトリーチ等の活動

年月日	名称	場所	参加人数	概要
2014年 11月7-9日	サイエンスアゴラ	日本科学未来館		研究成果の展示
2014年 3月1日	『聴覚障害者のための字幕付与技術』シンポジウム 2014	京都大学学術情報メディアセンター	119人	字幕付与技術に関する一般向けのシンポジウム http://www.ar.media.kyoto-u.ac.jp/jimaku/jimaku14.html
2013年 3月8日	『聴覚障害者のための字幕付与技術』シンポジウム 2013	京都大学学術情報メディアセンター	103人	字幕付与技術に関する一般向けのシンポジウム http://www.ar.media.kyoto-u.ac.jp/jimaku/jimaku13.html
2012年 4月1日・2日	CREST Symposium on Human-Harmonized Information Technology	京都大学百周年時計台記念館	132人	本CREST領域の3研究チームと共同で、世界のトップレベルの研究者と意見交換、及び本プロジェクトの広報 http://www.ar.media.kyoto-u.ac.jp/crest/sympo12/
2011年 10月1日	『聴覚障害者のための字幕付与技術』シンポジウム 2011	京都大学学術情報メディアセンター	67人	字幕付与技術に関する一般向けのシンポジウム http://www.ar.media.kyoto-u.ac.jp/jimaku/jimaku11.html
2010年 11月27日	『聴覚障害者のための字幕付与技術』シンポジウム 2010	京都大学学術情報メディアセンター	71人	字幕付与技術に関する一般向けのシンポジウム http://www.ar.media.kyoto-u.ac.jp/jimaku/jimaku10.html

§6 最後に

トップレベルのマルチモーダルな情報処理の実現を目指して、音響・音声・映像・言語処理の研究室が結集して研究を進めてきたが、試行錯誤の連続であった。主要なアプリケーションであるスマートポスターボードについては早くから設計を行ったが、当初は会話データの収録がなかなかうまくいかなかった。機材や担当者が多いため、丹念に準備をしても、いずれかのモダリティで欠損が生じたり、モダリティ間で時間の同期がとれなくなったり、後処理に大きな手間がかかったりして、完全なデータがなかなか得られなかった。経験・ノウハウを積み重ねた結果、研究後半にようやく効率的かつ効果的なデータ収録ができるようになった。この知見・ノウハウは貴重な財産と考えている。

システム構築においては、個々のメディアの処理の性能が最先端であるが故に処理能力の問題に直面した。対象人物が1名もしくはオフラインの処理は問題なくても、複数名をオンラインで処理するだけのシステムは現状の計算機資源では困難であった。クラウドサーバで処理を行うことも考えたが、音声・映像ともデータサイズが膨大なため伝送がボトルネックとなった。そのため、スマートポスターボード(液晶ディスプレイ)に設置するセンサと物理的に接続できる範囲で、GPU 付きのデスクサイドワークステーションで処理を行う構成となった。したがって、ポータブルなシステムの実用化には、まだ相当の計算機能力向上を待つ必要がある。

研究内容については、信号処理から興味・理解度レベルまで、音声と映像の統合を行ったので、興味深いテーマが次から次と出てきたし、国内外からの招待講演もこれまでになく多かった。今でもオンリーワンの研究であると考えている。

ICASSP 2012 におけるデモ展示にて