

私の音声研究と S L P

鹿野 清宏[†]

鹿野が行ってきた音声認識を中心とする音声・音響の研究を振り返るとともに、音声言語情報処理研究会（S L P）での活動についても述べる。

[†] 奈良先端科学技術大学院大学、名古屋大学

1972 年 3 月 名古屋大学大学院 工学研究科 修士課程を修了し、日本電信電話公社武蔵野電気通信研究所に入所し、基礎第四研究室、斎藤特別研究室で音声認識の研究に携わった。優秀な上司と同僚に恵まれ、非常に幸運でした。1981 年 7 月 名古屋大学から工学博士の学位を修得した。

電電公社から NTT の民営化の間隙に、1974 年からカーネギーメロン大学(CMU)で客員研究員として 2 年間滞在した。NTT での音声認識関連の研究成果をもとに、LPC 距離尺度、ベクトル量子化による音声認識の研究を進展させ、CMU の音声プロジェクトに貢献した。音声データベースに基づく統計的音声認識の重要性を認識して、日本の音声翻訳のプロジェクトを行う ATR 音声通訳研究所音声情報処理研究室に室長として、1986 年 6 月に CMU から直接大阪に着任した。

ATR では、音声認識研究の基礎となる大規模音声データベースの作成、HMM に基づく統計的音声認識の研究を立ち上げるとともに、コードブックマッピングによる話者適応化や声質変換、音声翻訳の研究、さらに、国内外の研究者を集め、視聴覚機構研究所聴覚研究室（東倉室長）と密接に連携して、ニューラルネットや特徴ベースの音声認識の研究を行い、国内外で非常に高く評価され、ATR の音声認識合成の基礎研究が世界で認知された。音声の研究とともに、若手研究者を中心とした勉強会などを通して、多くの音声研究者の育成に努めた。ニューラルネットによる音声認識(TDNN:Waibel et al.)では、IEEE Signal Processing Society, 1990 Senior Award(論文賞)を、大語彙連続音声認識システム(HMM-LR)の開発で、日本音響学会技術開発賞を受けた。

1990 年 2 月に NTT ヒューマンインタフェース研究所に戻り、音声情報処理グループリーダーとして、日本の第二世代携帯電話の音声符号化方式 VSELP の開発の指揮をとるとともに、大語彙連続音声認識の研究を行った。

1994 年 4 月に奈良先端科学技術大学院大学情報科学研究科 教授に着任し、音声情報処理と音情報処理の融合を目指して、音情報処理学講座を開設して、中村哲助教授と伊勢史郎助手とともに研究と教育を開始した。11 月には、音情報処理論の講義内容にもとにした教科書「音声・音情報のデジタル信号処理」を出版した。

また、学会でも、1995 年 11 月に情報処理学会音声言語

情報処理研究会(SLP)に大語彙連続音声認識研究データベース・ワーキンググループ(JNAS WG)を立ち上げ、19 大学、20 企業の協力を得て、新聞記事読み上げ音声コーパスの大規模音声データベース(306 人×503 文)を作成した。この音声データベース(JNAS)は、現在でも最も使われている音声データベースになっている。この JNAS は、音声データベース調査委員会(板橋先生主査)と共同で音響学会から配布された。この JNAS での大語彙連続音声認識アルゴリズムの評価ができるように、伊藤克亘先生をはじめとする JNAS WG の努力により、テストセットや標準の音韻モデル、言語モデルの構築を行い、大語彙連続音声認識などの研究開発に大いに役立った。さらに、学生や若手研究者の教育として、1996 年 8 月に SLP で音声認識基本技術講習会を、武田先生、河原先生をはじめとする WG メンバーのリーダーシップと杉山先生の協力で会津大学でサマースクールを開催して、大学や企業からの若手の研究者の育成を行った。以後、この講習会は現在まで、山形大学、京都大学と場所を変えながら、毎年開催され、音声研究者の育成に役立っている。

1997 年 4 月、この SLP のワーキンググループのメンバーを主体とした IPA 独創的先進的情報技術に係わる研究開発「日本語ディクテーション基本ソフトウェアの開発」が採択され、年間約 5 千万円の 3 年間のプロジェクトを開始し、大語彙連続音声認識ソフトウェア JULIUS(李、河原)、日本語言語モデル構築ツール Palmkit(伊藤彰則)をはじめ、多くの成果を上げた。さらに、音声認識基本技術講習会などにより、学生や多くの若手研究者の育成に努めた。

2000 年 4 月に、IPA のプロジェクトを進展させ、情報処理学会に連続音声認識コンソーシアム(CSRC)を 3 年間の期限で立ち上げた。情報処理学会では、初めてのコンソーシアムであり、特別幹事として、コンソーシアムの責任者となった。約 70 の企業・大学の参加を得て、会員へのセミナー、音声認識基本技術講習会、ソフトウェアやモデルの配布を毎年行った。大語彙連続音声認識プログラム Julius の改良、記述文法用認識エンジン(Julian)、音韻モデルとして、高精度成人モデル、高齢者モデル、子供モデル、電話用音韻モデルなどを、言語モデル作成ツールや言語モデル、Windows 対応などの開発を行い、会員に配布した。3 年目に、コンソーシアムの代表を河原達也先生に交替した。2001 年 5 月に音声認識基本技術講習会の資料をもとに、情報処理学会の IT TEXT として「音声認識システム」を出版

した。

音情報処理学講座では、中村助教授の活躍により、ハンズフリー音声認識やマルチモーダル音声認識の分野での研究も成果を上げた。また、伊勢助教の活躍により音場の再現の研究も進んだ。2000年3月に中村哲助教授がATRへ室長として転出し、同4月に猿渡洋助教授が着任しました。

2000年には、NEDOの高齢者音声認識のプロジェクトを高山サイエンスにラボを設置して研究を行い、収集した高齢者音声データベース(S-JNAS)(402人×200文)を公開した。この年から欧米と同様に、音情報処理学講座では、学生を研究員として雇用し、賃金を支払う教育研究体制を開始した。

2000年4月に猿渡洋助教授が、10月に李晃伸助手が着任し、新しい枠組みの研究、ブラインド音源分離(BSS)(猿渡、西川)、音声対話システムの研究を開始した。BSSでは、分離におけるひずみが小さいSIMO-ICAのアルゴリズム(高谷、猿渡)が見出された。BSSは、NTT-CS研との共同研究も行い、音源分離の世界の中心として認められるようになった。2000年度には、学生の西村の献身的な努力により、ロボティクス講座とロボット対話システムASKAを構築してデモを行った。2002年度、ASKAを発展させて生駒市北コミュニティセンターに音声案内システム「たけまるくん」(西村竜一)を設置し、実環境音声対話実験と実音声データの収録を開始した。さらに、無音声(NAM)認識・無音声電話(中島淑貴)や声質変換の研究が大いに進展した。2002年度から文部科学省の情報科学研究科のCOEのプログラムの中核メンバーとなり、5年間の研究が開始された。また、科研基盤(B)の二つの申請「人にやさしい音声認識」と「音響拡張現実感」が採択され、それぞれ4年間の研究も開始されて研究が進んだ。NEDOの高齢者ロボットコミュニケーションのプロジェクトでのハンズフリー音声認識の本格的な研究も始まった。

2003年度、文部科学省のリーディングプロジェクトe-Society基盤ソフトウェアの総合開発で「ユーザ負担のない話者・環境適応性を実現する自然な音声対話処理技術」が採択された。これまでのけいはんなの音声研究活動とSLPの大語彙連続音声認識の研究活動が評価されて採択されたと思われる。奈良先端大に京都大学、名古屋大学、和歌山大学、名古屋工業大学と企業3社を加えた年間約1億円の5年間の音声認識と合成のプロジェクトが開始された。

2005年度、SIMO-ICAのDSP実装が神戸製鋼との共同研究で進展して、NHKニュースなどマスコミで多く取り上げられた。戸田智基助手がCMUから着任し、音声モーフィングの研究が飛躍的に進展した。非可聴つぶやき声の認識(中島)で、電子情報通信学会の論文賞および最優秀論文賞「猪瀬賞」を授与された。次の年度でも、残響の復元(山丈、

猿渡)で電子情報通信学会から論文賞を受賞した。実時間BSS(森)は、ロボットの国際会議(IROS)でも高く評価され、論文賞を受賞した。これらの業績が高く評価され、2007年1月に「Contribution of speech recognition, dialog systems, voice conversion, and acoustic field realization」の業績でIEEE Fellow(米国電気電子学会フェロー)の称号が授与された。この年度には、大阪からの地下鉄が学研北生駒駅(奈良先端大学前)に開通したので、生駒市、近鉄、大学の要望に応じて、近鉄学研北生駒駅に音声情報案内システム「キタちゃん」とロボット案内システム「キタロボ」(川波、加藤、Cincarek等)を設置した。「たけまるくん」と同様、現在でも運用中であり、大量の音声データを収録している。

2007年度、文部科学省のリーディングプロジェクトe-Societyが音声対話システムの実環境運用、JuliusのSH-4マイコンでの実装(畑岡、小久保、李)、ブラインド音源分離、ハンズフリー音声認識BSSA(高橋)、非可聴つぶやき、音声モーフィング、STRAIGHT(河原英紀)、HTS(徳田)、発話障害者補助(戸田)など多くの成果を生み出して成功裏に終了した。このプロジェクトの成果をもとに、音声によるユニバーサルコミュニケーションの研究を発展させている。

騒音下でのロボットとのハンズフリー音声対話システムの構築を世界に先駆けて構築した。研究室の音声認識技術、ブラインド音源分離のBSSAとを融合した成果である。SIMO-ICAがブラインド音源分離の世界コンテストの非線形部門のトップになり、表彰を受けた。

2007年に文部科学省の基盤研究Aの「音声によるユニバーサルコミュニケーションの研究」が採択され、4年間の研究を開始した。2008年度、総務省の戦略的情報通信研究開発推進制度(SCOPE)に猿渡准教授と戸田助教がそれぞれ採択され、3年間の研究が開始された。文部科学省のリーディングプロジェクトe-Societyのプロジェクトの事後評価が行われ、非常に高く評価された。

2009年度、今までの研究室の研究成果が、国際的にも高く評価された年度であり、ICASSPでロボット対話、残響下での音声対話システム、声質変換のspecial sessionでの3件の講演が招待された。また、InterSpeechでも、戸田助教がSilent SpeechのSpecial Sessionで、NAM(Non-Audible Murmurと声質変換)でkeynote speechを行った。

2011年度、猿渡准教授のブラインド音源分離(BSS)とミュージカルノイズフリーの研究が高く評価され、猿渡准教授にドコモ・モバイル・サイエンス賞の基礎科学部門優秀賞や東北大学電気通信研究所のRICE Awardが授与された。さらに、高次統計量によるミュージカルノイズの削減アルゴリズムの研究も進展し、2013年に猿渡准教授に市村賞が授与された。

また、鹿野は、平成 2004 年 3 月に「大語彙連続音声認識の研究開発とその普及活動」の業績で情報処理学会からフェローの称号が、2004 年 9 月に「音声認識高度化と研究・教育」の業績で電子情報通信学会からフェローの称号が授与された。その他、情報処理学会から山下記念研究賞(平成 12 年)を、日本バーチャルリアリティ学会から論文賞(平成 13 年)を、IEEE/RSJ(IROS2005)で Best Application Paper Award(平成 18 年)を、IEEE MLSP2007 で Data Analysis Competition Winner(平成 19 年)を、電気通信普及財団からテレコムシステム技術賞(平成 21 年、24 年)などを授与された。

音情報処理学講座のスタッフ、学生も、多くの賞を受賞している。主な受賞は、スタッフでは、猿渡准教授が電子情報通信学会論文賞(平成 12 年、17 年)、ドコモ・モバイル・サイエンス賞(平成 23 年)、東北大学電気通信研究所 RIEC Award(平成 23 年)、日本音響学会技術開発賞(平成 21 年)を、戸田助教授(現本学准教授)が IEEE Signal Processing Society 2009 Young Author Best Paper Award(平成 22 年)、電子情報通信学会情報システムソサイエティ論文賞(平成 20 年)、エリクソン・ヤングサイエンティスト・アワード(平成 20 年)、日本音響学会の独創研究奨励賞板倉記念(平成 21 年)と栗屋潔学術奨励賞(平成 21 年)を、伊勢史郎助手(現、京大准教授)が日本音響学会論文賞(平成 10 年)を、李助教(現、名工大准教授)が日本音響学会の栗屋潔学術奨励賞(平成 14 年)、を情報処理学会の山下記念研究賞(平成 19 年)を受賞している。

学生のための主な受賞は、40 件あまりになり、主な受賞は、電気通信普及財団テレコムシステム技術学生賞(平成 9 年、12 年、15 年、16 年)、日本音響学会の栗屋潔学術奨励賞(平成 12 年、21 年)、C&C 若手優秀論文賞(平成 17 年)、船井情報科学奨励賞(平成 17 年)、日本音響学会の独創研究奨励賞板倉記念(平成 18 年、21 年)、エリクソン・ベスト・スチューデント・アワード(平成 24 年)などである。本学の NAIST ベストスチューデント賞は、音情報処理学講座で、2001 年の創設以来、ほぼ毎年受賞しており、修士課程が 8 回、博士課程が 6 回である。この 19 年間に、33 名の博士課程の学位修得者、約 200 名の修士課程の学生を育成した。

音情報処理学講座では、19 年間の活動において、多くのプロジェクトや企業との共同研究を行い、総計 10 億円に迫る外部資金を獲得するとともに、本学だけでなく、関西を中心とした音声情報処理、音情報処理の人材を育成してきた。19 年間の共同研究の延べ件数は約 60 件・約 1.4 億円、受託件数は約 70 件・約 4.9 億円、科学研究費は約 40 件・約 1.5 億円、奨学寄附金は約 60 件・約 0.6 億円である。

NTT、ATR、奈良先端科学技術大学院大学において、2013 年 3 月時点で著書 3 編、学術雑誌論文 178 編、国際会議論文

432 編等の多くの研究発表を行った。奈良先端科学技術大学院大学での研究は、多くの国からの資金や企業からの共同研究、研究プロジェクト分担に支えられてきました。

研究室が主体として行ったプロジェクトをあげておく。

- (1) 文部科学省 e-Society 基盤ソフトウェアプロジェクト 鹿野「ユーザ負担のない話者・環境適応性を実現する自然な音声対話処理技術」(2003 年度～2007 年度)(約 1 億円×5 年)
- (2) IPA に係わるプロジェクト
 - (a) IPA 独創的先進的情報技術に係わる研究開発 鹿野「日本語ディクテーション基本ソフトウェアの開発」(1997 年度～1999 年度)(約 5 千万円×3 年)
 - (b) IPA 未踏ソフト 戸田「固有声変換ソフトウェア」(2006 年度)
 - (c) IPA 未踏ユース 鎌土「オーディオオブジェクト操作」(2009 年度)
- (3) 文部省科学研究費
 - (a) 文部省科学研究費 基盤研究(C)「日本語音声ディクテーションシステム」(1996 年度～1997 年度) 鹿野
 - (b) 文部省科学科究費 基盤研究(C)「日本語音声ディクテーションシステムの環境への適応」(1998 年度～2001 年度) 鹿野
 - (c) 文部科学省科学研究費 基盤研究(B)「マイクロホンアレーを用いたハンズフリー音声認識アルゴリズム」(1999 年度～2002 年度) 中村(猿渡)
 - (d) 文部省科学科究費 奨励研究(A)「明瞭な自由発話を対象とした頑健な大語彙連続音声認識」(2001 年度～2002 年度) 李
 - (e) 文部科学省科学研究費 基盤研究(B)「環境、話者、タスクへの適応性をもつユーザにやさしい音声認識」(2003 年度～2006 年度) 鹿野
 - (f) 文部科学省科学研究費 基盤研究(B)「音情景の分解・合成・拡張に基づく音響拡張現実感」(2003 年度～2006 年度) 猿渡
 - (g) 文部科学省科学研究費 萌芽研究 「つぶやき音声による個人認証」(2005 年度～2006 年度) 鹿野
 - (h) 文部科学省科学研究費 若手研究(A)「任意のユーザを対象とする統計的声質変換・制御法」(2006 年度～2008 年度) 戸田
 - (i) 文部科学省科学研究費 基盤研究(A)「新しい音声メディアによるユニバーサルコミュニケーション」(2007 年度～2010 年度) 鹿野
 - (j) 文部科学省科学研究費 若手研究(A)「バリアフリー音声コミュニケーションのための次世代ボイスチェンジャー」(2010 年度～2014 年度) 戸田
 - (k) 文部科学省科学研究費 基盤研究(A)「高次統計量による自立カスタムメイド音コミュニケーション拡張

- システム」(2011年度～2014年度)猿渡
- (1) 文部科学省科学研究費 基盤研究(C)「機械学習および統計モデルに基づく音声対話システムの応答生成」(2012年度～2015年度)川波
- (4) 総務省SCOPEプロジェクト
- (a) 総務省 SCOPE 障害者の声質変換(2008年度～2010年度)戸田
- (b) 総務省 SCOPE 補聴器(2008年度～2010年度)猿渡
- (5) 情報処理学会コンソーシアム 鹿野「大語彙連続音声認識フリーソフトウェア」(2000年度～2003年度)
- さらに、分担として行ったプロジェクトをあげておく。
- (1) 文部科学省科学研究費
- (a) 文部省重点領域研究「音声・言語・概念の統合処理による対話の理解と生成に関する研究」(1994年度～1995年度)堂下
- (b) 文部省総合研究(A)「音の知覚と情報統合に関する総合研究」(1995年度～1996年度)寛
- (c) 文部科学省科学研究費特定領域(A)「英語学習」(2000年度～2002年度)中川
- (d) 文部科学省科学研究費 学術創成「ソフトウェアロボット」(2001年度～2003年度)田中
- (e) 文部科学省科学研究費 基盤研究(A)「音韻モデルの分散学習」(2003年度～2005年度)武田
- (f) 文部科学省科学研究費 基盤研究(B)「デジタルメディアコンテンツ作成のための多様な音声の合成技術」(2005年度～2008年度)徳田
- (g) 文部科学省科学研究費 基盤研究(C)「多元観測信号を用いた音信号の予測及び復元」(2006年度～2008年度)板倉
- (h) 文部科学省科学研究費 基盤研究(A)「聴覚・音声機能の支援・拡張技術」(2007年度～2010年度)河原英紀
- (2) 科学技術振興事業団 CREST
- (a) 科学技術振興事業団「脳をつくる」「聴覚の情景分析に基づく音声・音響処理システム」(1997年度～2002年度)河原英紀
- (b) 科学技術振興事業団「高度メディア社会の生活情報技術」「日常生活を拡張する着用型情報パートナーの開発」(2000年度～2005年度)木戸出
- (c) 科学技術振興事業団「高度メディア社会の生活情報技術」「The Processing of Expressive Speech」(2000年度～2004年度)Campbell
- (d) 科学技術振興事業団「共生社会に向けた人間調和型情報技術の構築」「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」(2009年度～2012年度)河原達也

- (3) 文部科学省「21世紀COEプログラム」「ユビキタス統合メディアコンピューティング」(2002年度～2006年度)千原
- (4) 経済産業省 NEDOプロジェクト
- (a) 「シニア支援システム」「音声インタフェース」(2000年度)釜江
- (b) 「NEDO戦略的先端ロボット要素技術開発プロジェクト」(2006年度～2008年度)小笠原
- (5) 総務省SCOPEプロジェクト
- (a) 総務省 SCOPE 発話障害者(2005年度～2007年度)平原
- (b) 総務省 SCOPE NAMマイク(2008年度～2009年度)平原
- (6) IPA 独創的先進的情報技術に係わる研究開発「擬人化エージェント」(2000年度～2002年度)嵯峨山
- (7) 文部科学省 都市エリア 胎児心音の検出(2008年度～2009年度)
- (8) RWC 音響データベース(1997年度～2000年度)

また、音情報処理学講座は、多くの企業から共同研究や受託研究などを通してお世話になりました。企業との共同研究などをまとめておく。

トヨタ自動車(2004-2012)、ヤマハ(2007-2012)、サクサシステムエンジニアリング(2012)、フォスター電機(2009-2012)、パナソニック(1995, 2000-2004, 2007-2011)、NEC(1995-2000, 2006-2010)、ソニー(2007-2009)、ホシデン(2008-2009)、三洋半導体(2008)、生駒市(2002-2008)、日立(2002-2007)、旭化成(1994, 2002-2007)、メガチップス(2006)、オムロン(1998-2005)、KDDI(2005-2006)、神戸製鋼(1998, 2003-2006)、デンソー(2004)、NTT(1994-1995, 1997, 2001-2003)、日産自動車(2001-2004)、松下電工(2001-2004)、P to P(2002)、熊本ソフトウェア(2002)、ハーネス総合技術研究所(1999-2000)、アドバンスドメディア(1999)、東芝(1997-1999)、ATR(1994-1998)、NTTデータ(1995-1997)、音響測器(1994-1996)、NTTアドバンスドテクノロジー(1996)、ブラザー工業(1996)

この40年あまりの研究を通して、多少なりとも、世界の音声研究の発展に貢献できたと思います。また、多くの研究者、学生の育成に貢献できたことは、大きな喜びです。

音情報処理学講座

<http://spalab.naist.jp/>

鹿野のページ

<http://spalab.naist.jp/staff/teacher/shikano/shikano.html>

音情報処理学講座の研究室紹介(論文リストなど)

http://spalab.naist.jp/config/config_list.html