

人間と機械が対話する夢を追って 音声認識研究・開発を支える共通基盤を開発

英語で「ディクテーション」といえば、「口述筆記」のことだが、知識情報処理の世界では「大語彙連続音声認識」を意味する。これは、人が口述したことを、コンピュータに正確に文書入力する技術であり、人間と機械の対話を実現するのに不可欠な技術の一つだ。この研究分野に、IPAの「イノベ創的情報技術育成事業」で開発されたソフトウェアがある。京都大学学術情報メディアセンターの河原達也教授らが開発した「日本語ディクテーション基本ソフトウェア」だ。

このソフトウェアは、フリーウェアとして公開され、国内外の大学や企業の研究所で、研究・開発用プラットフォームとして広く利用されている。

「あ、い、う、え、お」から始まった音声認識

日本の音声認識研究の歴史は、一九五〇年代にまで遡ることができる。まさにコンピュータの創生期と重なっている。ソビエト連邦がスプートニクを打ち上げ、米ソの宇宙開発競争の扉を開いた頃、京都大学の坂井利之教授らは音声認識装置の開発に苦闘していた。計算に不可欠なメモリが極めて貧弱だった頃であり、わが国においてコンピュータがまだ誕生して間もないころであった。

一九五九年、坂井氏らは苦勞して世界で初となる「音声タイプライタ」を開発した。音声タイプライタは、研究室の壁を覆いつくし、まるで洋服ダンスをずらりと並べたように、とてつもなく巨大な装置だった。第一号機は真空管方式だった。音声で「あ、い、う、え、お」と入力すると、タイプライタが「ア、イ、ウ、エ、オ」と打った。短母音しか入力できなかつた。それでも見学者からは、驚嘆の声が上がった。音声認識における日本の研究水準の高さを示す快挙だった。

坂井氏のもとで研究を支えた若き研究者がいた。当時、大学院生だった堂下修司氏（現・京都市立大学名誉教授）である。堂下氏こそ、日本の知識情報処理研究の草分けであり、音声メディア処理や人工知能などの分野で数多くの先駆的な業績を挙げた研究者だ。（写真1）音声タイプライタは、今日の音声認識システムの源流となっている。

一九八七年、すでに情報工学科教授であった堂下氏の研究室に、一人の大学院生が配属された。「日本語ディクテーション基本ソフトウェア」を開発した河原達也氏である。（写真2）河原氏は、学部生の時代に坂井研究室で文字認識と図形認識を研究し、堂下教授のもとで音声の研究することになった。

* 1 創的情報技術育成事業：大学・国立研究所・企業等に存在している有望なソフトウェア技術のシーズを発掘し、産・学・国立研究所の協力のもとに、その成果を公開して自由な利用を図ることによって、ソフトウェア技術全般の向上に寄与することを旨とした事業。

「当時、堂下研究室では、自然言語の論理研究やシステムの定性推論など、人工知能的なテーマに積極的に取り組んでいた。音声研究は最もマイナーなテーマだった。私も人工知能に関するテーマを希望したのだが、堂下先生の勧めで音声をやることになった。ただし『単なる音声屋になるな』というのが先生の口癖だった。以来、私は十数年、音声認識研究に集中してきた。音声認識は実に奥深く、かつ面白いテーマだった。」

河原氏が音声研究に着手した一九八〇年代後半から、音声認識研究は世界的に活発化し、その技術は一気に進展していったのである。

加速した日米の音声認識研究

音声認識の歴史を振り返ると、古くは一九五一年にベル研究所が行った「ゼロ交差数」方式を用いた数字の音声認識への取り組みが挙げられる。音声は波形で表すことができるが、ゼロ交差数とは、上下に振れる波形がゼロ値と交差する数であり、この数から波形の周波数の近似値を得て、特徴量を求める方式である。この特徴から人が何と発声したかを判定するのである。

実用化につながった技術革新としては、一九七〇年代にロシアと日本から同時に提案された「DPマッチング法」や、日本単独提案の連続した数字を



写真2 京都大学 河原教授



写真1 京都大学坂井研究室で開発された「音声タイプライタ（第2号機）」
向かっているのは1961年当時の堂下名誉教授。

出典：京都大学堂下修司先生退官記念会 1999年6月発行 堂下先生退官記念集「知」

認識できる「2段DPマッチング法」がある。DPマッチング法とは、音声の各音韻の発声時間の伸縮に注目した方法である。例えば、「発声」という言葉なら「は・っ・せ・い」の各音韻で発声時間が異なり、言葉によって時間が伸縮する。この伸縮を正規化しておき、最も似たパターンを求める。パターンのマッチング方式として動的計画法（ダイナミック・プログラミング:DP）を用いることからDPマッチング法と呼ばれる。

その後、米国では、統計的手法を用いて音素単位の特徴量パターンで音響的にモデル化する「HMM^{*2}」を用いた研究が盛んになる。HMMに代表される統計確率手法は、今日の音声認識の標準的手法となった。

このように音声認識は、膨大な音声パターンと言語パターンの統計データの集積によって実現される。言語パターンとは、単語間の接続関係を規定したもので、「文法モデル」や、ある単語の次にくる単語の出現確率のかたよりに見る「統計モデル」が用いられている。

河原氏は、一九八〇年代後半から一九九〇年代前半にかけて、米国防総省高等研究計画局（DARPA^{*3}）が実施した一連のプロジェクトによって、音声認識研究のスピードが加速したと指摘する。

「米国は、DARPA主導で大規模なデータベースを整備し、この共通の土俵の上で『競争と協調』の原理で研究を進めていった。私は、こうしたやり方に冷やかな見方をして

いたところがあったし、単にデータを増やすのは芸（アイデア）がないと思っていた。この認識の違いが、とりわけ大語彙連続音声認識（ディクテーション）の研究では米国が一步前進することになった。データベースの規模を大きくすることに、それほど意義を感じていなかったのだが、実際に収集データのスケールが大きくなると質的にも変わった。」

DARPAは、音声認識の応用イメージを示し、それを実現するために複数の研究機関に資金を提供し、その成果を競わせた。その結果、認識性能を短期間に向上させることに成功したのであった。単語ごとにスペースを入れて記述する英語と異なり、より高度な形態素解析が必要な日本語のディクテーションは難易度が高い。しかし、河原氏らは、「日本語ディクテーション基本ソフトウェア」の開発において、米国が果たした認識性能レベルに追いつき、追い越そうと考えたのだった。

大学や研究機関の壁を越えたコラボレーション

このような状況において、河原氏（当時は助教）らが、IPAの「独創的情報技術育成事業」の公募に「日本語ディクテーション基本ソフトウェア」の開発を申請し、採択されたのは、一九九七年四月のことである。三年間に亘る研究開発の体制は、研究統括として奈良先端科学技術大学院大学から鹿野清宏教授、音声認識グループとして河原氏、音韻モデルグループとして名古屋大学から武田一哉教授が参加。さらに言語モデルグループと

* 2 HMM: Hidden Markov Model: 隠れマルコフモデル
* 3 DARPA: Defence Advanced Research Projects Agency

して電子技術総合研究所から伊藤克亘氏（現在、名古屋大学助教授）、読み付与グループとして京都高度技術研究所から山田篤氏らが加わり、強力な布陣となった。河原氏は、「二三年間に亘る研究プロジェクトを開始するに際して、あらかじめ年度毎に数値目標を設定しておき、これを着実に達成していく必要があった。音声認識の研究では、そうした数値目標の設定には、豊富な研究経験が必要となる。そこで、鹿野先生が研究統括としてリーダーシップを発揮された。鹿野先生抜きには、このプロジェクトはありえなかった」と語る。

組織の壁を超えたプロジェクトになる兆しはすでであった。一九九五年頃には、伊藤氏がモデルに必要なデータを集めようと呼びかけた。そして、武田氏もソフトウェア基盤の必要性を呼びかけた。この呼びかけによって、情報処理学会のワーキング・グループでは毎日新聞の記事七年分のテキストデータが収集され、日本音響学会の委員会には三〇六人が新聞記事一五〇文を個々に読み上げた音声データが集まった。こうして、ディクテーションに必要な高精度な日本語言語モデルと日本語音韻モデルを開発する下地は整った。

一般に、大語彙連続音声認識システムを実現するためには、精度の高い音響モデルと言語モデル、そして効率の良い認識エンジンが必要となるが、このような高度なシステム開発と要素技術の研究をバランスよく推進していくには、共通のプラットフォームが必要となる。河原氏らが開発した日本語ディクテーション基本ソフトウェアは、まさにその共通

プラットフォームとなっている。同ソフトウェアは、標準的な音声認識エンジン、日本語音響モデル、日本語言語モデル、そして日本語形態素解析／読み付与ツールなどから構成される（図1）。それぞれが独立したモジュールとなっており、各モジュールのフォーマットとインタフェースには一般性があるため、開発者の手で目的に応じて作成・編集・置換したり、認識プログラムの設定を自由に変更することができる。つまり、同ソフトウェアは汎用性と一般性の特徴とする。

河原氏らのグループが開発した認識エンジンは、「Julius」（ジュリウス）と命名された。プロジェクト最終年度のJuliusの認識率は、ほぼリアルタイム（実時間の二倍以下）で処理する高速版で九二%、高精度版では九五・八%に達し、画期的な認識率を示した。

プロジェクトの最終年度、河原氏はシアトルにあるマイクロソフト社を視察した。

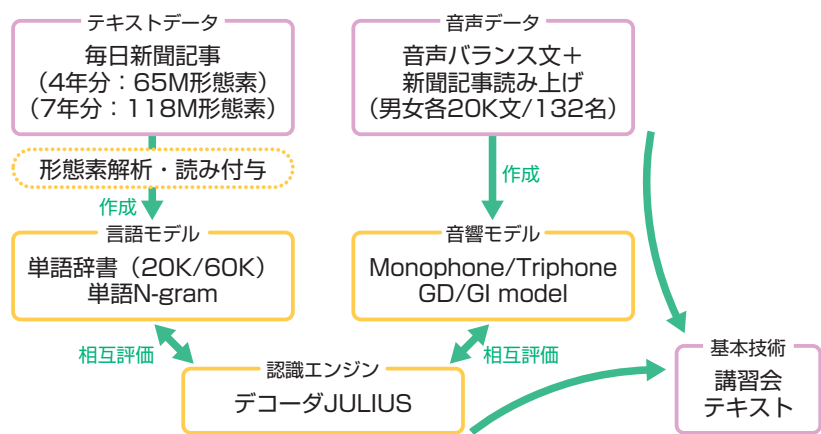


図1 日本語ディクテーション基本ソフトウェアのシステム構成
提供：京都大学 河原教授

「彼らに日本語ディクテーション基本ソフトウェアを披露した。皆、我々のソフトウェアに驚いていた。その時点で、我々の大語彙連続音声認識の研究が、世界水準に到達したと感じた。」

IPAからの受託事務をはじめ会議室の提供など、このプロジェクトを側面からサポートし続けた財団法人京都高度技術研究所 事業担当部長 三好則夫氏は、当時の様子を次のように振り返った（写真3）。

「鹿野先生、河原先生、武田先生。この三人のコンビが実に素晴らしかった。会議では、ざつくばらんに議論がかわされた。一年目、二年目、三年目と、年々認識率は良くなった。評価では、当日の新聞記事を読み上げて認識させた。認識性能が上がっていく様子を見るのが、私の楽しみの一つとなった。このプロジェクトを通じて、それぞれの大学や研究機関が持つ優れた知見を共有化することができた。河原先生も、論文に表れない情報を得られたとおっしゃっていた。三年間に及んだプロジェクトの最大の成果は、そこにもあるかもしれない。」

三好氏は、プロジェクトの終了後、このソフトウェアを何とか普及させたい、という河原氏らの意向を汲み取り、資金獲得に動いた。しかし、開発支援の枠組みとしてはIPAの「独創的情報技術育成事業」があっただ

の、残念ながら同ソフトウェアの普及に適した支援の枠組みがなかったために断念することになった。

そこで、情報処理学会の音声言語情報処理研究会内に「連続音声認識コンソーシアム」を設立し、音声認識を用いたアプリケーション開発を考える企業や研究機関から会費を集めることによって、同ソフトウェアの普及促進、改良・メンテナンスを継続的に行っていくことにした。同コンソーシアムには、主要な研究開発機関のほとんどが賛同し、企業三六社、一七の大学が会員として参加した。京都高度技術研究所は、同コンソーシアムの運営も支えてきた。

成功、そして次のステップへ

国内で音声認識の研究・開発に携わる人で、日本語ディクテーション基本ソフトウェアを知らぬ人はいないだろう。研究論文の引用数の多さがそれを裏付けている。国内ばかりではなく海外の研究者らも利用している。河原氏らの狙い通り、研究・開発用の共通プラットフォームとして普及し、成功した。さらに一歩進めて、実用化を視野に、河原氏を直接訪ねる人もいて、デモも用意されている。（写真4）

河原氏は、「国会関係者が来られたことがある。議事録の作成を音声認識でやれないかと。すでに話者認識機能（話者が変わったことを認識し、テキストに話者が誰であるかを示す



写真3 (財) 京都高度技術研究所 三好氏

タグを付けられる)も実装済みであり、国会での討論への対応を進めている」と語る。奈良県の「生駒市北」コミュニティセンター—STA はばたき」には、同ソフトウェアを使用した音声情報案内システム「たけまるくん」が稼働している。同センターの施設案内、同市の観光案内、同センター周辺の情報案内など、各種案内を行うシステムである。来館者の質問に、アニメーションが答える仕組みだ。

日本音響学会の春季／秋季研究発表会(全国大会)で発表されている音声認識関連の論文において同ソフトウェアの利用状況は、音声認識関係の全論文数の二〇%弱に達する。五つの研究のうち二つは、必ず同ソフトウェアを利用している計算になる。同ソフトウェアを用いた研究開発の成果は、様々な分野で利用されつつある。カーナビの音声によるコマンド制御、医療レポート等の入力支援システム、介護や福祉分野の音声リモコン、教育分野では語学学習支



写真4 日本語ディクテーション基本ソフトウェアのデモ風景

援システム、インターネットの音声ポータルサイトの構築、放送関連では聴覚障害者のための字幕スーパ、携帯電話やPDA(携帯情報端末)などのモバイル機器の音声入力、さらには自動翻訳サービスや会話ロボットなど、実に幅広い。

河原氏の研究室ではマイクロソフト社および東京大学との共同研究も行われている。^{(*)4}オンラインヘルプに音声認識を利用する研究だ。ユーザはパソコンに接続したマイクロホンで、知りたいことや質問を音声で入力すると、その質問がブラウザにテキスト入力される。(写真5)ユーザの質問に対して、オンラインヘルプが音声でこたえる。このような研究は、米国リハビリテーション法五〇八条が影響しているといえるだろう。

国内でも二〇〇〇年六月に通商産業省から「障害者・高齢者等情報処理機器アクセシビリティ指針」が出されており、音声認識に言及している。情報バリアフリーの実現手段として、音声認識はその重要度をますます高めている。

河原氏はい、「不特定話者(あらかじめ話者の声を登録し学習させる必要がない)の連続音声認識の基盤はできた。ただ、騒音環境下での音声認識や『話し言葉』の認識については課題がある。また、人間と音声で対話する機械、『音声対話システム』の実現については、これまで実用化に至ったものはほとんどない。まさにSF映画の世界だが、音声対話システムについては、まだ解決しなければならない研究課題も数多くある。しかし、この分野での日本の研究の蓄積は相当大きい。様々な研究が各組織で深く静かに進められてい

* 4 オンラインヘルプ (online help) : ユーザの利用を手助けするための助言をコンピュータ画面で閲覧できる仕組み。必ずしもネットワークで利用するものではない。

C O L U M N

画像処理サブルーチン・パッケージ「SPIDER」シリーズ開発小史 日本の画像処理研究・実利用化を手助けした「道具箱」

学会の研究活動とその実利用化に寄与した、
もう一つの事例

先の「日本語ディクテーション基本ソフトウェア」は、近年の音声認識研究の分野において、学会の研究活動とその実利用化に寄与した好例といえるだろう。実は、過去においても、同じ様にIPAのサポートで開発されたソフトウェアが学会の研究活動とその実利用化に大きく貢献した事例がある。それが画像処理サブルーチン・パッケージ「SPIDER II」だ。ジャンルは異なるが、間違いなく画像処理研究・開発を支え、日本の研究・技術水準を引き上げたソフトウェアだ。「SPIDER」シリーズとして研

究者や企業の技術者たちの間で広く利用された、このソフトウェアの歴史を振り返ってみよう。

パターン認識・処理の研究が始まってほぼ三十年が過ぎ、文字・図形・物体・音声・自然言語処理など、それぞれの分野ごとに様々な成果が生まれ、実用レベルに達するものが出てきている。画像認識・処理分野では、まさに「目に見える」成果が多岐にわたって生まれている。パソコンには「お絵かきソフト」から高級な画像加工ソフトまでがそろい、スキヤナやデジタルが使われ、産業・工業分野では、衛星画像のリモートセンシング、顕微鏡画像や医療診断用CTスキヤンの解析、非破壊検査システム、視覚センサなどが活躍している。指紋や虹彩を識

るが、ときには研究者らが組織の壁を越えて取り組んでいくべきテーマだと思っ。」

「音声認識のメリットを多くの人々に」という、河原氏の音声研究への意欲は尽きない。



写真5 マイクロソフト社および東京大学との共同研究「ダイアログナビ」

別するバイオメトリクス（生体認証）分野もある。カメラ付き携帯電話がスキヤナ代わりになるもの：いや、もうきりがない。この分野は十分に成果をあげたのではないのか、とさえ思う。

しかし、ある画像処理分野の研究者がこういう。

「人間が一目見れば直ちに分かるのに、コンピュータにやらせるとできなかつたり、たとえても時間がかかる問題は、まだまだ沢山あります。これはコンピュータによる人間のパターン認識のゴールがまだまだ遠いということ。認識研究と処理技術の両方が刺激し合わなければ研究は進歩しないのです。」この研究者とは、芝浦工業大



写真1 芝浦工業大学大学院 高木教授

学大学院教授・東京大学名誉教授の高木幹雄氏（写真1）である。

画像処理アルゴリズムとサブルーチンとは？

リモートセンシングや非破壊検査の権威でもある高木氏からは、こう教えてもらった。

「日本の画像処理技術は世界でも最高水準にあります。それは研究の重要な段階に、優秀な道具^{*}に恵まれたからです。「SPIDER」および「SPIDER II」というサブルーチン・パッケージです。」

「SPIDER」および「SPIDER II」（写真2）とは何か？
 キャッチフレーズは「画像処理サブルーチン・パッケージ」とある。英文名称から直訳すれば「画像データの



写真2 「SPIDER II」とマニュアル

強調と認識に役立つ作業命令文集」となる。「画像処理に使う基本的なアルゴリズムを集大成したもの」とも説明される。

アルゴリズムとは算法と訳される。簡単なたとえでいえば、「コーヒーをいつもおいしく入れる手順」もまたアルゴリズムである。私たちが日常的に、デジタル画像を元にして、タテ長の顔にしたり、縁取りを強調したり、明るさの補正などを行っている。こうした処理の手順それぞれがアルゴリズムであり、コンピュータには、「デジタル・アルバムソフト」などの中で動くサブルーチンとして組み込まれている。簡単な手順の後ろに、一連の処理プログラムが動いている。それに気づくことは少ないが、さらに後ろには、永年にわたる研究者たちの英知が集約されているのだ。

サブルーチン間の連携は、例えば顕微鏡で見た金属加工部品の表面積を測定したいという場合、加工の結果生じた「模様の違いの検出」、

「輪郭の抽出」、「輪郭の平滑化」、「面積測定」という順にプログラムが読み出される。

「研究者の誰もが知っていた」サブルーチン集

ひとくちに画像処理というが、その研究分野は①データの圧縮・符号化技術、②画質改善・強調・復元技術、③CTスキャンのような画像の再構成技術、④画像の特徴を抽出する認識・理解技術の四つに大別できる。「SPIDER」および「SPIDER II」は、一九八〇年代初頭から半ばにかけて世界中で公表されたこれらの分野の、基本的に有用なサブルーチンをひとまとめにし、研究者・技術者向けに配布したものである。「SPIDER」（四四二本）は一九八二年に、その補強版「SPIDER II」（三五八本）は一九八六年に販売されている（括弧内はサブルーチン数）。驚くのはその販売実績で、前者は約千セット、後者は約二百セットに

C O L U M N

「このプロジェクトを最初から見守ってきた高木氏の言葉で、締めくくろう。」

「プロジェクト最初の委員会からもう二十年以上も経ってしまったけれど、その価値は今も失っていません。彼らの論文や著作も含めて、この分野の『古典』になった。今の学生たちもC言語に移植したものをパソコンで使って勉強しているでしょう。日本の画像処理研究をこれから牽引してゆくのは彼らです。」

「画像処理の分野は産業的応用が広がり、次々と多様化していますが、後にCTなどに応用された『再構成』技術のようにエポックになる研究は少なくなりました。まだまだ、より深く研究し

たこと」も挙げている。開発者たちのスタンスの取り方、気配りと粘り強い努力が、見事な結果につながったハッピーな例といえる。

これからも若い技術者たちの支援を

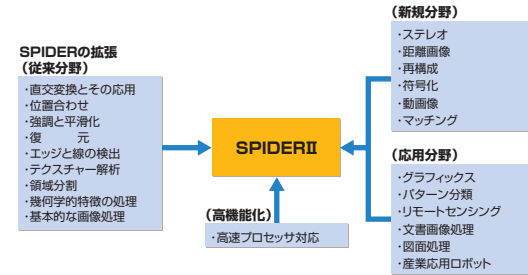


図1 SPIDER II への変化
出典：「SPIDER II ユーザーズ・マニュアル」をもとに作成

のぼる。公的研究機関や関連企業の数からすると、「この分野の研究者なら誰でも知っていた」（高木氏）というのもうなずける。では、誰が作ったのか？「SPIDER」は、通産省工業技術院・電子技術総合研究所（現、独立行政法人産業技術総合研究所）の田村秀行氏らが、当時、東京大学生産技術研究所にいた高木氏らの協力を得、また名古屋大学の鳥脇純一郎教授らの先駆的研究の助けも借りて開発したものである。国の機関が開発して成果を公開したソフトの第1号となった。

この四年後に世に出る「SPIDER II」（図1）は、IPAの一八四年度特定プログラム開発支援制度を利用して、委員会形式（高木幹雄委員長）で、前出の研究者を主体に開発したものである。

「SPIDER」シリーズが広く世に受け入れられた理由について、高木・鳥脇・田村氏らは、著書の中で、「移植性の高さ」と低価格にあった」と書いている。当時、研究者の間で使われていたプログラミング言語「FORTRAN」で書かれ、中規模以上のコンピュータなどの機種でも動くようにした配慮が実ったのだ。また、他の理由として、「かなり充実した利用者マニュアルとその副産物として詳細なサーベイ（論文調査）が公表されていた」。

なくてはならない領域が残っていることも忘れてはならない。その意味で、IPAのような機関が、若い研究者・技術者たちから生まれてくる本質的な研究に目を行き届かせて支援していただきたいものだと思います。」

「SPIDER」シリーズは、当時の画像処理に関するアルゴリズムを網羅することによって、日本の画像処理技術の水準を底上げし、その研究開発の裾野を拡大した。世界最高レベルに至った日本の画像処理技術の歴史は、まさに「SPIDER」シリーズなくして語れないのである。

* 2 SPIDERプロジェクト開発リーダー。立命館大学理工学部情報学科教授。日本バーチャルリアリティ学会複合現実感研究委員会委員長なども務める。

* 3 名古屋大学の鳥脇教授（現中京大学教授）らが、同様のサブルーチン・ライブラリ「SLIP (Subroutine Library for Image Processing)」を配布していた。「SPIDER開発に対して全ソースリストの提供と、整理・統合のための仕様変換に協力していただいた」（田村氏）。

* 4 別冊 O plus E、画像処理アルゴリズムの最新動向、高木・鳥脇・田村共編、新技術コミュニケーションズ社、1986年11月初版