

音声言語をキーワードとした 40 余年の研学生活

中川聖一

概要: 音声言語情報処理研究会(SLP)が発足してから 20 年が経過した。その直前の 5 年間の準備期間を含めると 4 半世紀を共に歩んできたことになる。SLP 創設前後の状況と私自身の音声言語処理に関する 40 余年間の研学生活を振り返り、最後に今後の展望を述べる。

キーワード: 音声言語情報処理研究会、電子情報通信学会第 2 種研究会、情報処理学会研究グループ、HMM、ニューラルネットワーク

Research Life of 40 and More Years on Spoken Language Processing

Seiichi Nakagawa

Abstract: Twenty years have run after founding the Special Interest Group on Spoken Language Processing. Before founding, we had five years as a preparatory period. After all, I have always been a friend of this community during a quarter of a century. I look back before and after the foundation, and also 40 and more years on spoken language processing for my research life. Finally I describe the perspective in future.

Keywords: Spoken Language Processing, Research Group in IPJ. Research Group in IEICE, HMM, Neural Network

1. はじめに

人間は、思いのほか適応性があり、自分の歩んできた道を肯定的に捉える性癖があるように思う(悲観主義者も多いが、悲観ばかりしていると生きていけない)。私の研学生活でも、振り返ってみれば、各エポック点で「運が良かった」と感じる事が多くある(「幸運」と「幸福」は異なるらしい)。

一つ目は、大学院の学生・助手時代に研究環境が恵まれた点が挙げられる。恩師の坂井利之教授(文化功労者)の研究室は、情報処理分野で我が国は言うにおよばず世界を牽引していて、日本をリードしていくことが当たり前のような感覚が身に沁み込んだことである。

二つ目は、新構想の大学院大学である豊橋技術科学大学に奉職したことである(この選択が良かったと言うことにしておく)。当時の大学院進学率が国立大学工学部

で 20%~50%程度の時代に、全員が大学院に進学するというシステムは、研究に恵まれた環境であったと思う。

三つ目は、若手教員(35 歳以下)を対象に文部省在外研究員制度が発足し、1 年遡って募集することになり、当時 36 歳であった私が全国の 40 人の枠に選ばれたことである。

四つ目は、この在外研究員の長期出張の帰国後、電子情報通信学会の出版委員であった千葉成美氏から音声認識の専門書の執筆を依頼されたことである。「確率モデルによる音声認識」[1]を執筆し、お陰様で高い評価を得た(坂井先生から最初の本は良い本を出すと言われていた)。今から思えば、本のネーミングも良かったと思う。

五つ目は、本研究会の発足に繋がる一連の活動が、次節で述べるように、皆様からの温かい支持を得たことである。電子情報通信学会の第 2 種研究会(1990-1991)、情報処理学会の研究グループ(1992-1993)を経て、1994 年 4 月に念願の音声言語情報処理研究会を発足させることができ、今日に至っている。

六つ目は、1990 年に 41 歳で教授に昇進し、43 歳頃

1 豊橋技術科学大学 情報知能工学系
Toyohashi University of Technology
nakagawa@tut.jp

から大講座(教授3人、助教授3人、助手3人、技官1人)の最年長になり(一人が副学長に移動、一人が他大学に転出のため)、人事等を一手に引き受けてきたことである。一線級の人を採用しようと心がけてきたが、有能な方の転出等に出くわすなどストレスもあったが、この20年以上、他人からの圧迫感はほとんどなかった(ほとんどないという意味は、説教されたことが2回ある(誰からとはここでは言わない))。

七つ目は、今後の音声分野の方向はドキュメント処理と、皆様からの協力を得て音声ドキュメント処理ワークショップを開催し、実行委員長として、今年度で8回目を迎えたことである。

この間、科学研究費・特定領域研究の計画班班長を務めたり、種々のプロジェクトに加えていただき、良い経験をさせていただいた。大学では、系長(学科長)や各種センター長、教授会議長、JABEEプログラム責任者なども歴任したが、これらは流れでやってきただけという気がする。

本稿では、SLP研究会創設前後の活動と私の研究で受賞した4つの論文の紹介を行い、40余年間の研究生生活を振り返り、最後に今後の展望を述べる。

2. 音声言語情報処理研究会(SLP)の発足前後の状況

SLPの前身は、電子情報通信学会第2種研究会「マルコフモデルとニューラルネットワークを包含する新しい音声認識手法時限研究会」と情報処理学会「音声言語情報処理と音声入出力装置研究グループ」に遡る。この発足の経緯は文献[2][3]に詳しく述べているが、簡単に紹介しておく。

2.1 電子情報通信学会第2種研究会(1990-1991)

1989年4月より、文部省科学研究費総合研究「マルコフモデルとニューラルネットワークを包含する新しい音声認識手法の総合的研究」(代表:中川聖一)が採択され(3年間)、比較的狭い範囲で参加者を募り、泊まり込みで議論を始めた(毎回20-30名前後の参加で、一人1時間の持ち時間で徹底的な討論を行った)。その様子を見られていた藤崎博也先生や白井克彦先生から、公開討論の場にした方がよいとの助言を受け、電子情報通信学会の第2種研究会「マルコフモデルとニューラルネットワークを包含する新しい音声認識手法時限研究会」を2年間の時限で設立した(1990-1991;委員長:中川聖一、幹事:小林豊先生、嵯峨山茂樹先生)。24年前の研究会名であるが、現在でも立派に通用する名称であるのは感慨深い。研究会は、年2回温泉地で合宿形式で開催したことから通称「音泉研究会」が始まった(あと2回は、総合研究会の会合)。この合宿形式の研究会は当時としては珍らしく、夜の2次会で本音の情報を得たり親睦を深めることに寄与した。第1回の研究会(1990年7月)では、パネル討論「音声認識研究—これからの10年」が開催され、その書き起こしを元に、32回SLP研究会(2000年7月)に、「10年前のパネル討論の検証」として、報告した[4]。以下、4回の開催状況を示す

(3回のパネル討論の書き起こしは、<http://www.slp.cs.tut.ac.jp/>の研究領域のページを参照)。

第1回:「音声認識研究—パイオニア研究者からの提言とこれからの10年」湯河原温泉厚生年金会館(1990.7.20-21)、参加者数62名。

第2回:「隠れマルコフモデル・ニューラルネットワークシンポジウム」箱根静雲荘(1991.2.8-9)、参加者数58名

第3回:「音声認識の新潮流をめざして」蒲郡形原温泉鈴岡(1991.7.12-13)、参加者数62名

第4回:「連続音声認識シンポジウム」館山寺温泉ビューホテル開華亭(1992.2.7-8)、参加者数72名。

2.2 情報処理学会研究グループ(1992-1993)

2年間の活動を経て、情報処理学会に音声関連の研究会がないことに従来から不満を感じていたことから、情報処理学会に音声関連の研究会設立を目標に、先ず、「音声言語情報処理と音声入出力装置研究グループ」を1992年に新設した(1992-1993;主査:中川聖一、幹事:岡田美智男氏、小林豊先生、新田恒雄氏)。年3回の合宿形式で開催した。なお、電子情報通信学会の第2種研究会は「音声認識の実用化をめざす新手法時限研究会」(新美康永委員長)と名称を変え、完全な共催で研究会を開催した。

第1回:「音声言語処理と対話理解」上諏訪温泉湖泉荘(1992.7.13-14)、参加者数85名。

第2回:「音声入出力機器の現状とヒューマンインタフェースの課題」塩原簡易保養センター(1992.10.23-24)、参加者数50名。

第3回:「音声入出力機器の動作環境に関するシンポジウム」箱根静雲荘(1993.2.5-6)、参加者数37名。

第4回:「Spontaneous Speechの分析・理解・生成」熱川ハイツ(1993.7.2-3)、参加者数48名。

第5回:「音声入出力機器の現状とヒューマンインタフェースの課題シンポジウム」熱海KKR熱海竜泉閣(1993.10.22-23)、参加者数45名。

第6回:「音声認識・理解の今後の課題」蒲郡西浦温泉シーサイド南風荘(1994.2.4-5)、参加者数41名。

2.3 情報処理学会研究会(1994-)

2年間の研究グループの活動を経て、研究会の設立が無事承認され、1994年4月に音声言語情報処理研究会が発足した(主査:中川聖一、幹事:岡田美智男氏、小林豊先生、新田恒雄氏・1996年4月からの幹事:岡田美智男氏、小林哲則先生、畑岡信夫氏)。栄えある第1回の1番目の発表は藤崎先生にお願いした(「音声言語処理の理念と課題」)。4番目の発表は、今でも語り草となっている嵯峨山先生の「なぜ音声認識は使われないのか・どうすれば使われるのか?」であり、当時の熱気が懐かしい。私自身は、最後の7番目に「音声言語情報処理の研究課題」を発表した。研究会の位置づけと研究方向を図1と図2に再掲しておく[3]。

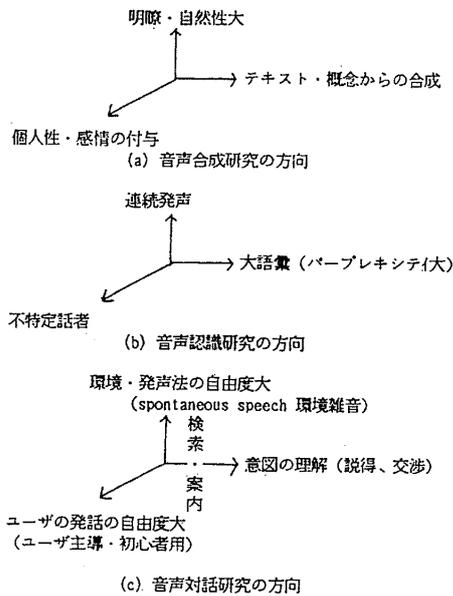


図1 音声研究の方向

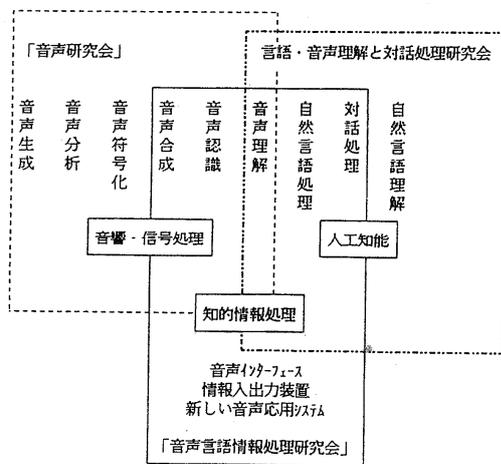


図2 関連する3研究会の位置づけ

私の主査の間(1994-1997)の簡単な活動記録は、情報処理学会誌の7月号に毎年掲載されている。1泊2日の音泉研究会の参加者数(年2回)は、順に、45名、36名、71名、66名、66名、77名、59名、59名と60名前後で推移した。この4年間は音声対話システムやマルチモーダル対話システムが盛んであったと記憶している。いずれの開催も印象に残っているが、特に印象深いのは、第6回研究会を新宿の工学院大学で開催した「話し言葉の文法構築は可能か」という音声と言語の両分野の研究者に関心の高いパネル討論を行い、参加者数

が約80名であったことである。

その他の主査としての活動として、

- ・「大語彙連続音声認識研究用ワーキンググループ」の立ち上げ(1995年11月)
- ・情報処理学会誌特集号「音声言語情報処理の現状と研究課題」(1995年11月号)
- ・第1回「日本語ディクテーション基本技術講習会」(1997年8月)
- ・情報処理学会論文誌特集号「音声言語情報処理」(1999年4月号)
- ・情報処理学会英文図書“Spoken Language Systems”IOS Press(2005年)

などがある。英文図書の発行は、計画から7、8年を要し、皆様に大変ご迷惑をお掛けした(岡田美智男氏と河原達也先生に大変お世話になった)。

3 論文賞になった論文紹介

3.1 A Speech Understanding System of Simple Japanese Sentences in a Task Domain[5]

音声理解システムと大袈裟なタイトルを付けたが、本研究期間は、アメリカで音声理解プロジェクトが開始されており、それを意識してタイトルにした。この研究は、私の博士論文の中核で、以下の提案を行った。

- (1) 並列木探索法： 当時の人工知能分野では、木探索法としては、A*探索、best-first探索、breadth-first探索が中心であったが、best-firstとbreadth-firstの折衷である、一定数の枝を同時に伸ばしていく木探索法を提案した(1974年)。この伸ばし方は、単語同期であったが、フレーム同期のものが**ビームサーチ**と呼ばれているものである(Lowerre:1976)。
- (2) 直接マッチング法： 発話文の音声認識結果の音素系列から、音素列で表現される単語を検索する方法を提案した(1975)。これは、短い区間と長い区間とのマッチング結果を平等に比較できる非対称型DPマッチング法である。これをフレーム単位で行う方法が**連続DPマッチング法**である(最初に提案された連続DPマッチング法は対称型であり、研究会で誤りを指摘したことがある。岡：1978)。
- (3) 文脈自由文法制御型連続音声認識法： 当時、ネットワーク文法と称する正規文法制御型連続音声認識法は研究されていたが、文脈自由文法(Earley法によるトップダウン)制御の連続音声認識法は世界で初めてである。後に、係受け制御型連続音声認識法(尾崎：1986)、文脈自由文法のCKYボトムアップ制御型連続音声認識法(Ney:1987)が開発された。なお、この両者が同一に定式化出来ることを示した(中川：1987)。
- (4) MAP学習による話者適応： 単語中の母音部を同定し、この母音のスペクトルを用いて、個人差のMAP話者適応を行う方法を提案した(1976)。音声認識にMAP学習法を用いた世界で最初の研究だと思われる(但し、平均ベクトルだけの学習)。

3.2 Speaker-Independent English Consonant and Japanese Word Recognition by a Stochastic

Dynamic Time Warping Method[6]

私が、カーネギーメロン大学長期出張中(1985.3-1986.2)の最初の4ヶ月で行った研究(英語音素認識)を、帰国後、日本語の単語認識にも適用し、それらをまとめた論文である。長期出張の前の1983年頃から、ベル研究所のグループがHMMによる音声認識の研究論文をBSTJに発表し、一気にHMMが注目され出した。その当時は、離散出力分布型HMMが研究の中心であり、この欠点を感じていたので、連続出力分布型HMMとDTWの統合を、長期出張中の研究課題の一つにした(もう一つの研究課題は、博士論文の文脈自由文法制御型連続音声認識法を、フレームレベルで実現したもの: Computer Speech and Language, 1989)。この研究のことを、CMUの訪問に来られた斎藤収三先生と古井先生にお話ししたことを覚えている。なお、連続出力分布型HMMの最初の発表は1985年だと思われる(Rabiner, Juang: 1985)。

3.3 音声認識研究の動向[7]

電子情報通信学会論文誌に、サーベイ論文というジャンルが設けられ、牧野正三先生から音声認識のサーベイを依頼され、論文としてまとめたものである。1988年に出版した「確率モデルによる音声認識」以降の研究動向と自論を展開した。最新の研究動向を、音声分析・特徴パラメータ、音響モデル、言語モデル、認識システムに分けて詳述した。また、自論として、音声認識とテキスト処理における言語モデル・言語処理の役割の違いを指摘し、アプリケーションが定まれば、トライグラム言語モデルは人間の言語処理モデルと遜色がないこと、音響モデルは人間の音韻知覚能力に遠くおよばないことを、定性的・定量的に示した(図3参照。なお、これを充実させたものは、Computer Speech and Languageに掲載予定、2014)。このサーベイ論文は、2000年度と2001年度の電子情報通信学会論文誌アクセス数1位にランクされた。

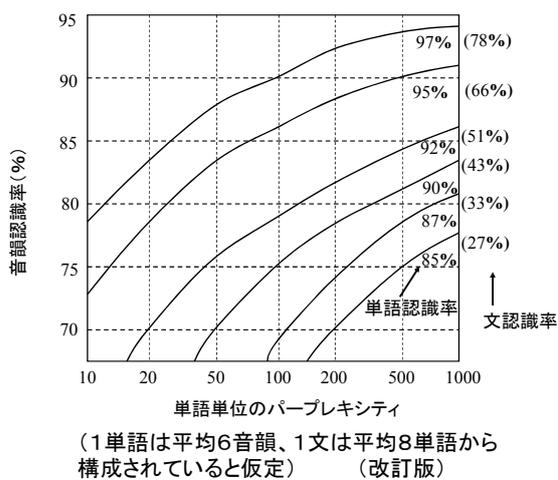


図3 音韻認識率(言語情報なし)、単語認識率、パープレキシティの関係

3.4 Hidden Conditional Neural Fields for Continuous Phoneme Recognition[8]

この論文は、当時博士課程の学生であった藤井康寿君(現・Google)の博士研究の3つ目の論文である。藤井君は、大語彙連続音声認識システムSPOJUSの継ぎ接ぎだらけのプログラム(約20000行)をC++に書き換え(SPOJUS++)、最後には、DNN-HMMとWFSTも研究の手段として実装してくれた。いつかのSLP音泉研究会で、藤井君が「世界の音声認識システムを創る」と宣言したことを覚えておられる方もいると思うが、本当にそれを実現しようと頑張った成果の論文である。DNNとHMMを包含し、音響モデルと言語モデルを同時に識別学習できる枠組みが評価されたものと考えている。この研究の継続・発展が私の残された研究課題の一つである。

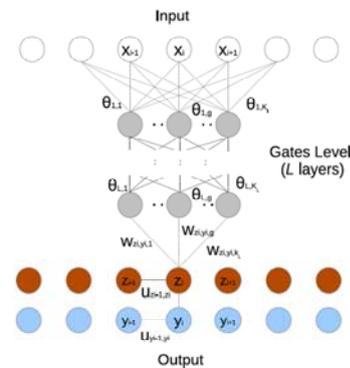


図4 Deep-HCNFの構造

4 ニューラルネットワークによる音声処理の再考

最近、再びニューラルネットワークによる音声処理に注目が集まっている。これは、Hintonらによる多層ニューラルネットワークの学習法(制約付きボルツマンマシンによる事前学習)が発明されたことによる。原理的に万能といわれるニューラルネットワークの能力を引き出せるようになったからで、1990年前後から最近までの動向は、文献[9]で紹介した。一方、1990年前後には、既に多くのアイデアや問題点などが議論されていた。当時の私自身の考えをまとめた文献[10]から、論点を再掲し、現時点の考察をもとに補足をする。囲み文章が文献[10]からの引用である。

1. HMMとニューラルネットワークの長所と短所

- ##### 1.1 共通点
- 大量のデータに基づく統計的学習法である
 - 哲学「量的拡大は質的变化をもたらす」
 - 知識工学「知識は力なり」
 - パターン認識「量は質を凌駕する」
 - 学習のための量の問題(データ量、計算量など)
 - 認識モードは学習モードに比べて非常に早い。
 - 規則型知識の組み込みが難しい

- ・理論物理学よりも実験物理学の立場（データ駆動型の意味で）

[補足 1.1] 学習のための量の問題は、計算機環境の発展で当時の 1000 倍の能力向上で、ほぼ解決したに近い。

1.2 HMM の利点と欠点

- ・音声の生成過程は確率過程であるか（LPC 分析、HMM）？
- ・静的パターン認識としては、連続出力分布モデルではベイズの識別規則による 2 次判別関数である。

[補足 1.2] 生成モデルである HMM による音声合成は、この時点では思いつかなかった。

1 次と 2 次回帰係数で、ほぼ連続関数が近似できる点が味噌。パラメータの識別学習に言及していないのも私のミス。

1.3 ニューラルネットワーク（多層パーセプトロン型）の利点と欠点

- ・「ブラックボックス」、「無手勝流」であり明示的な知識が得られていない場合は便利・有用である。
- ・静的パターン認識に対しての識別能力は高い（区分的非線形判別関数）。マルチプレート・LVQ に匹敵する。パターン認識の本質は、パターンマッチングであり、研究は明白な設計方針、モデルのもとにコンパクトに標準パターンをどう構成するかどうマッチングするか歴史であった。
- ・時系列パターンの処理に弱い。
- ・知識の組み込みや処理の制御が難しい。
- ・出力値の物理的意味がない。（ON/OFF の 2 値情報）
- ・ → 学習法の改善で解決可能？
- ・新しいカテゴリの追加に対し、再学習が必要である

[補足 1.3] 最近の研究で、リカレントニューラルネットワークにより、時系列処理にも十分対処できることが示されている。出力値の物理的意味は、事後確率であり、これは当時から既知であったが、出力ラベルの与え方が 2 値という指摘であったと思われる。

3. ニューラルネットワークの疑問と課題

3.1 ニューラルネットワークは HMM を包含するより一般的な枠組であるのか

- ・HMM はニューラルネットワークの一分野であるとの立場
- ・HMM は LVQ よりも神経回路網モデルに近い。
- ・原理的に何でもできるものは特殊化された問題にも強力か。
チューリングマシン V.S. 有限オートマトン/プッシュダウンオートマトン
アセンブリ言語 V.S. LISP/Prolog/Fortran/C

[補足 3.1] 有用な処理機能は、遺伝的に組みこまれているという観点からの考察が必要。

3.2 ニューラルネットワークは人間の神経回路網のモデルになりえているか。

- ・脳の神経細胞（ニューロン）数は約 100 億個、Fan-in、Fan-out 数は約 1000~10000 個である。これと比べ

て現在のモデルは「太平洋の海水容量をバケツで計量する」ようなもの

- ・ネコは人間の音声言語を獲得・学習できない。ネコの生理的解明で何が得られるか。
- ・物質の性質・組成を調べるのに、分子レベル/元素レベル/原子レベル/素粒子レベルのいずれが適切か？（弁別特徴/音素/音節/形態素/単語？）
- ・工学的モデル・認知心理学モデルと割り切るべきか。

[補足 3.2] 最近のニューラルネットワークの規模は、隠れ層 7 層、各層のユニット数 2000 程度でも十分学習できる。これはパラメータ数にして、約 3×10^7 であり、画像処理用のネットワークの規模はもっと大きい、バケツがプールくらいにはなっている。

3.3 ニューラルネットワークで新しいルール型知識が獲得できるのか。

- ・わかっていることしか得られない。では検証モデルか？知識が分散表現の場合、説明できるのか（逆問題？）
- ・知識組込みの明白なモデルとの比較によってのみ、ニューラルネットワークの知識獲得の有効性が検証できるとするなら無用。

[補足 3.3] コンピュータ将棋は、膨大な棋譜からパターンを学習し、プロ棋士を凌駕するまでになった。また、クイズ番組では、膨大な知識を記憶・推論し、人間のチャンピオンを負かした。しかし、言語の自動獲得は依然難しい。パターン認識技術の更なる向上が不可欠である。

3.4 パターン処理と記号処理の融合は可能か？

- ・推論、大部分の確立されたアルゴリズムなど記号処理の方が得意な分野がある。
- ・ニューラルネットワークによる構文解析・有限オートマトンの実現の研究例はあるが不完全である。
- ・人間は厳密なアルゴリズムを実行しておらず大雑把であり、ニューラルネットワークとよく似ているが、人間の方が個々の精度ははるかに優れているので大雑把でもうまくいっている。
- ・一つの良い例が言語の確率モデル（確率文法、trigram、共起頻度の利用）である。
- ・現在の最も優れた連続音声認識手法は、音素・音節などの基本単位モデル（HMM など）の連結（記号処理）と入力音声全体との照合（パターン処理）であり、中間結果（音素・音節ラティス、単語ラティス）を介さない手法である。

[補足 3.4] HMM や言語モデルのパラメータの推定値の誤差が数%でも認識率に影響するが、ニューラルネットワークによる言語モデルの有効性が実証されているので、ニューラルネットワークによる数値表現能力は高いが、ニューロンでも同じか疑問。最後のコメントは、WFST に帰結した。

4. 結言

4.1 何故ニューラルネットワークか

- ・ニューラルネットワークは面白い。
- ・ニューラルネットワークが得意とする分野があるので

従来手法にうまく組み込むのがよい。

- ・連続音声認識に適用するには、出力値の物理的意味付けが必要。たとえば確率的尺度に変換する。
- ・ニューラルネットワークに時系列処理をもたせればもたせるほど HMM に似てくる。HMM は非常によいモデルである。→この結論では進展がない。

【補足 4.1】 ニューラルネットワークの出力の物理的意味は、事後確率と分かっているが、クラス帰属確率への変換のことを指している。ニューラルネットワーク(DNN)と HMM を包含する手法として、我々は、DNN-HCNF を提案している。

4.2 音声認識の進展に寄与したトピックス

- ・1960年代 電子計算機によるシミュレーション
- ・1970年代 LPC 分析、DP マッチング
- ・1980年代 HMM、スペクトル距離尺度
- ・1990年代? データベース、ニューラルネットワーク、聴覚・知覚モデル、言語処理、??

【補足 4.2】

1990年代の??の部分は、識別学習。

2000年代 WFST、HMM 音声合成、オープンソース、技術・手法の統合化の時代

2010年代 本格的実用化、ニューラルネットワークブームの再来

最後に、昨年7月に音声研究会で講演した際[9]に行ったアンケート結果を報告しておく。

質問「人間の脳処理とニューラルネットワークは関係あるか?」

- ・ほとんど関係ない (約3人)
- ・深い関係がある (約6人)
- ・関係あると思うが、ひよっとすると関係ないかもしれない (約25人)
- ・関係ないと思うが、ひよっとすると関係あるかもしれない (約20人)

5. おわりに

私の研究の主な成果は、電子情報通信学会音声研究会の招待講演「音声言語処理研究の進展と今後」(2006年12月)で、主な研究100として列挙した[11]。その後の主な研究30を列挙しておく。1パスデコーダSPOJUS(2006)、講義音声データベースと分析(2007)、インドネシア語のQAシステム(2007)、フィルターの挿入モデル(2007)、位相情報を用いた話者認識(2007)、単語単位の音響モデル(2007)、長短窓のケプストラムによる正規化(2007)、多チャンネルLMSによる残響処理(2007)、分散マイクロフォンアレイによる音源同定と発声方向推定(2008)、英語音声の明瞭度推定(2008)、音声のプライバシー保護(2008)、VAD信頼度のデコーダへの組込(2008)、長時間位相スペクトルの音声認識への応用(2009)、ポーズの挿入モデル(2009)、nグラムインデックスによる高速音声ターム検索(2009)、トピック依存言語モデル(2009)、構音障害者の音響パラメータと印象評価の相関(2009)、複数の対話エージェントとの対話システム(2009)、SPOJUS++(2010)、ソフトクラスタリング

による不特定話者の音声認識(2010)、講義音声の自動整形と可読性の向上(2010)、隠れニューラル場による音声認識(2010)、クラス言語モデルによる未知語の言語モデル(2011)、NMFとVQによる音楽重畳音声の分離(2011)、疑似ピッチ周期に基づく位相情報による話者認識(2011)、整形・非整形部分の自動検出(2012)、オンライン発音評価・明瞭度評価(2012)、ダミー音節を含む音節トライグラムによる音声ターム検索(2012)、音声ドキュメント内の人名抽出(2013)、講義音声の字幕表示法の比較(2013)、音節単位DNN-HMM&WFST(2013)。

今後の高齢・少子化・福祉時代および癒しに重きを置く時代に突入するに従い、音声処理技術がどのような役割を果たしていくか、また、音声処理技術が広く普及するに従い、プライバシー処理をどうするか[12]が重要になってくると思われる。

参考文献

1. 中川聖一：確率モデルによる音声認識、電子情報通信学会、1988
2. 中川聖一：「電子情報通信学会第2種研究会発足にあたって」時限研究専門委員会資料、SPREC-90-1, pp.1-11 (1990.7)
3. 中川聖一：音声言語情報処理の研究課題、音声言語情報処理、SLP1-7 (1994.5.20)
4. 中川聖一：10年前のパネル討論の検証—1990年代の音声言語処理研究を振り返る—、音声言語情報処理、SLP32-5(2000.7)
5. T. Sakai and S. Nakagawa: A Speech Understanding System of Simple Japanese Sentences in a Task Domain, IECCE Trans., Vol. E60, No.1, pp.13-20 (1977).
6. S. Nakagawa and H. Nakanishi: Speaker-Independent English Consonant and Japanese Word Recognition by a Stochastic Dynamic Time Warping Method, IETE Trans., Vol. 34, No. 1, pp.87-95 (1988)
7. 中川聖一：音声認識研究の動向、電子情報通信学会論文誌、Vol. 83-DII, No.2, pp.433-457 (2000).
8. Y. Fujii, K. Yamamoto and S. Nakagawa: Hidden Conditional Neural Fields for Continuous Phoneme Recognition, IEICE Trans. Vol. E95-D, No. 8, pp. 2094-2104 (2012)
9. 中川聖一：再訪：ニューラルネットワークによる音声処理、電子情報通信学会、音声研究会、SP2013-59(2013.7)
10. 中川聖一：HMMとニューラルネットワークの接点、科研費総合研究(A)「マルコフモデル・ニューラルネットワークを包含する新しい音声認識手法の総合的研究」第2回研究報告会資料(1989.7.14)
11. 中川聖一：音声言語処理研究の進展と今後、情報処理学会、音声言語情報処理、SLP-64(5)、電子情報通信学会、音声研究会、SLP2006-89 (2006.12)
12. 中川聖一、山本一公、土屋雅稔：音声に含まれるプライバシー情報の保護、人工知能学会誌、Vol.24, No.2, pp.190-195 (2009.3)