

TR-I-0309

## 3年間の音声情報処理研究をふりかえって

音声情報処理研究および研究支援業務概観

杉山 雅英

1993年3月5日

### 内容梗概

本資料では自動翻訳電話研究所 音声情報処理研究室において筆者の行なってきた研究、発表論文一覧および研究支援のための種々の業務に関して概説し、今後の課題などを述べる。

© ATR Interpreting Telephony Research Labs.

© ATR 自動翻訳電話研究所

# 3年間の音声情報処理研究をふりかえって

音声情報処理研究および研究支援業務概観

杉山 雅英

ATR 自動翻訳電話研究所

## 目次

1	はじめに	1
2	研究概観	1
2.1	Neural Network による連続音声認識の研究	1
2.1.1	Neural Network による音響セグメント話者適応	1
2.1.2	教師信号の自動生成による構造の学習	1
2.1.3	雑音抑圧ニューラルネットの音声認識での評価	1
2.1.4	NN と HMM との融合	1
2.1.5	識別誤り最小規準による話者適応 NN の最適化	1
2.1.6	FPM-LR による連続音声認識と追加学習の効果	2
2.1.7	FPM-LR による不特定話者連続音声認識	2
2.1.8	ニューラルネットワークを用いた話者独立の特徴抽出	2
2.1.9	NN による大語彙単語音声認識のための予備選択	2
2.2	知識処理と HMM による音素ラベリングの研究	2
2.2.1	音素ラベリングシステム	2
2.2.2	ラベリングワークベンチ	3
2.3	雑音環境下での音声認識の研究	3
2.3.1	コードブックマッピングによる雑音下音声認識	3
2.3.2	雑音環境下における HMM と TDNN の文節認識性能の評価	3
2.3.3	セグメント特徴による耐雑音 HMM	3
2.3.4	レベル可変雑音モデルによる雑音抑圧法	3
2.3.5	識別誤り規準を用いた耐雑音 HMM	3
2.3.6	標準 HMM による電話音声認識	3
2.4	音響特徴量による言語識別の研究	4
2.4.1	20ヶ国語の言語識別	4
2.4.2	Bilingual 音声を対象とした日英言語識別	4
2.4.3	スペクトル遷移の確率モデルによる言語識別	4
2.5	信号源モデルによる音声の区分化・識別問題の研究	4
2.5.1	話者特徴に基づく音声区分化と識別	4
2.5.2	男女性差に基づく音声区分化と識別	4
2.6	その他	4
2.6.1	歪み尺度測地線を用いた音声スペクトルの補間	4
2.6.2	伝送誤りに強い VQ 符号帳構成法	5
2.6.3	LPC ケプストラム係数の存在領域の諸性質	5
3	発表論文題目一覧	6

4	研究支援業務概観	14
4.1	研究室運営業務	14
4.1.1	研究室 Regular Talk の運営	14
4.1.2	Final Talk, 実習生報告会の運営	14
4.1.3	各種研究室日程の作成と室員への周知	14
4.1.4	アルバイト学生の確保	14
4.1.5	実習生の確保	14
4.1.6	ATR Working Group の主催	15
4.2	研究所運営業務	15
4.3	研究室 計算機及びソフトウェアの整備	15
4.3.1	計算機環境の整備	15
4.3.2	DEC 定例打ち合せ	15
4.3.3	デモプログラムの作成支援	15
4.3.4	外部ソフトハウスによる計算機ソフトの作成	16
4.3.5	計算機ファイルバックアップ業務	16
4.3.6	デモプログラムの解説書作成	16
4.3.7	ソフトウェア解説書作成	16
4.3.8	TeXstyle file の整備	16
4.3.9	文献データベースの作成整備	16
4.4	講演会、分担執筆及びシンポジウムの運営	17
4.4.1	技術講演の講師	17
4.4.2	ATR テクノロジーシリーズの執筆分担	17
4.4.3	第2種研究会 “連続音声認識シンポジウム” の運営	17
4.5	音声研究会幹事	17
4.5.1	ATR での音声研究会の主催	17
4.5.2	“ビギナーのための音声処理ワークショップ” の企画	18
4.5.3	“国際音声処理ワークショップ (IWSP93)” の企画	18
4.5.4	「不特定話者音声認識」特集号の企画	18
5	むすび	18

## 1 はじめに

本報告書は筆者が ATR 自動翻訳電話研究所 音声情報処理研究室に在籍した間に行なった研究の概観、発表論文の題目一覧及び研究支援のための種々の業務の概観である。

研究項目を以下に述べる。

1. Neural Network による連続音声認識の研究
2. 知識処理と HMM による音素ラベリングの研究
3. 雑音環境下での音声認識の研究
4. 音響特徴量による言語識別の研究
5. 信号源モデルによる音声の区分化・識別問題の研究
6. その他

## 2 研究概観

### 2.1 Neural Network による連続音声認識の研究

#### 2.1.1 Neural Network による音響セグメント話者適応

話者適応は教師ありと教師なしとに分類される。教師なしの方法は発声者に対する負担が小さいという点で優れているが、一般的には教師ありの方法に比べて、性能が低い。教師ありの方法に関して VQ codebook 間の写像構成に基づく方法が研究されてきた。さらにこれを NN で実現する方法が提案されその有効性が示されている。その発展として時間構造を持った音響セグメントの間の写像構成問題に取り組み、恒等写像による初期学習と未知発声者の単語音声による学習とを組合せた新たな方式を提案し、その有効性を示した。[5][12][21][45] [59][62]

#### 2.1.2 教師信号の自動生成による構造の学習

音声の諸分野において2つの集合の要素の対応付けに帰着される問題は数多い。集合の各要素の対応付けに関する何らかの教師信号に基づいた写像構成方法は強力であるが、教師信号を与えなければならないという点で柔軟性に乏しい。NN の重み係数を教師なしで自動的に学習できれば、より多くの学習パターンを提示できることになり、未学習パターンに対する識別性能を向上できる可能性を持つ。音素認識実験を通してその有効性が示した。[3][13][92][39]

#### 2.1.3 雑音抑圧ニューラルネットの音声認識での評価

雑音抑圧への応用は NN による非線形写像の1つの例としてはやくから研究された。入力に雑音下の音声、出力に雑音除去音声を提示することにより、写像を構成する。雑音抑圧の性能を被験者による聞き取り実験によって評価している。その結果、従来の spectral subtraction 法よりも優れていることが示された。現時点で問題になる処理量の多さに関しては、専用の並列演算装置を用いることで解決されるであろう。また、認識においては、波形上での写像ではなく特徴量上での写像を構成することにより、高速化が図れるものと思われる。VQ codebook マッピングによる方法との比較検討もなされており、それらの組合せにより性能が向上することを得た。[4]

#### 2.1.4 NN と HMM との融合

音声認識性能の向上を目的とした NN (Neural Network) と HMM とを融合する試みがいくつかなされている。本研究では TDNN (Time Delay NN) の出力の時系列を離散 HMM で表現する TDNN-HMM 認識方式を提案し、音素認識においてその有効性を評価した。[10][71][87][88]

#### 2.1.5 識別誤り最小規準による話者適応 NN の最適化

近年、ニューラルネットワーク (NN と略す) が音声認識などの諸分野で活発に研究されている。これらの NN は主に識別への応用であり、著者等は NN を用いた話者間の写像の検討を行ないその有効性を明らかにしてきた。一方、認識モデルのパラメータの最適化の手法として、識別誤り最小規準の最適化手法が提案されその有効性が検討されている。本研究では話者写像ニューラルネットワークの識別誤り最小化規準を用いた最適化の手法について述べた。[19][65][78][95]

### 2.1.6 FPM-LR による連続音声認識と追加学習の効果

ファジィパーティションモデル (以下 FPM) とその追加学習による連続音声認識について述べる。FPM は多入力多出力素子を持つニューラルネットワークである。FPM の出力値は非負でその総和が 1 であるため、入力音声の各音素に対応して確率的になる。FPM を LR パーサと結合し (以下 FPM-LR)、連続音声認識実験を行った。認識性能は、時間遅れニューラルネットワーク-LR (以下 TDNN-LR) と比較した。また、FPM とビタビアラインメントとを結合した自動追加学習を行い、連続音声への適応を行った。追加学習処理により音声データは、セグメンテーションされ FPM の追加学習に用いられる。実験結果から FPM が TDNN よりも高い認識率をもつこと、さらに、自動追加学習が FPM に対してより有利に働くことも示した。[22][47][64][77] [40]

### 2.1.7 FPM-LR による不特定話者連続音声認識

NN(Neural Network) を用いた不特定話者音声認識として、既に多数話者学習による TDNN を用いた連続音声認識に関して報告した。しかしながら、不特定話者の音素識別を TDNN に学習させるためには、多数話者の多量のサンプルを用いる必要があり、学習に時間がかかるという問題がある。一方、NN モデルとして FPM(Fuzzy Partition Model) を用いた FPM-LR による連続音声認識が提案され、既に特定話者に対する有効性が報告されている。FPM は短時間に学習が可能という特徴を持っており、本稿では FPM のこの特徴に着目し、音素識別に FPM を適用した不特定話者連続認識について評価した結果を報告する。FPM-LR を用いた不特定話者連続音声認識の評価を行なった。FPM は学習時間が TDNN の 2 分の 1 以下で済み認識性能の面でも TDNN を上回っていること、複数の FPM-LR を用いることで不特定話者に対する認識性能の向上が図れることが示され、文節認識率 80.0% が達成された。[29][51] [66][81][101]

### 2.1.8 ニューラルネットワークを用いた話者独立の特徴抽出

不特定話者に対する連続音声音声認識は、ニューラルネットワークを用いて実現できることが既に報告されている。不特定話者ニューラルネットワークは一種の不特定話者の特徴抽出器であるので、不特定話者に対する特徴抽出もニューラルネットワークによって実現の可能性がある。離散 HMM に対しては不特定話者のコードブック作成法が提案されている。本研究では、不特定話者の認識に向けて、話者ごとに異なる特徴をニューラルネットワークを用いて話者正規化する方法を提案する。話者正規化と同時に、任意の次元数の特徴をスペクトルから抽出することができる。また、このニューラルネットワークを HMM などと結合することにより不特定話者に対する認識精度の向上を期待できる。[30][52][74]

### 2.1.9 NN による大語彙単語音声認識のための予備選択

大語彙単語音声認識における処理時間短縮の為、単語予備選択の研究が行なわれている。本稿では、ニューラルネットワークの発火パターンを単語予備選択に利用する方法を提案する。TDNN(Time-Delay Neural Network) は単語中の音素識別及び単語音声認識において高い性能を示すことが報告されている。しかしながら、TDNN-LR による大語彙単語認識は処理に時間がかかるという問題がある。本研究では TDNN の発火パターンを単語予備選択に利用し、音素スキミングの以後の処理時間の短縮に関して検討を行なった。ニューラルネットワークの発火パターンを利用した単語予備選択方式を提案し評価した結果、処理時間の短縮が計られることが示された。今後の課題としては、 $F_p$  の分散を考慮する効果や  $V_w$  の初期クラスタリングの効果による予備選択の圧縮率の向上と棄却率の減少に関する検討、音素スキミングの処理時間の短縮、不特定話者に対する評価、更に大語彙単語での評価が挙げられる。[34]

## 2.2 知識処理と HMM による音素ラベリングの研究

### 2.2.1 音素ラベリングシステム

音声の自動ラベリングシステムの構築を目指し、まず、HMM を用いた音素セグメンテーション能力の評価を行った。さらに、スペクトログラムリーディング知識と HMM に基づくハイブリッド構成の音素セグメンテーションシステムを提案し、セグメンテーション性能評価を行い、高精度な音素セグメンテーションが実行できることを示した。[8][15][35] [60][63]

## 2.2.2 ラベリングワークベンチ

音素ラベルデータの効率的作成・保守のための支援ツールであり、EWS上で稼働する音素ラベリングワークベンチを開発した。ワークベンチは、操作性の良いユーザーインタフェース機能と自動ラベリング機能を備えている。[31][54][68][104][107]

## 2.3 雑音環境下での音声認識の研究

### 2.3.1 コードブックマッピングによる雑音下音声認識

雑音環境下音声認識における音声認識の問題を、雑音環境下音声空間から無雑音音声空間への写像構成問題として捉えた雑音環境下音声認識手法を提案した。両空間の写像を実現する手段として、コードブックマッピング手法を用いた。[43][57]

### 2.3.2 雑音環境下における HMM と TDNN の文節認識性能の評価

TDNN と HMM を識別器に選び、雑音環境下での HMM-LR および TDNN-LR の性能評価と、認識性能の改善法を述べた。HMM-LR では、無音区間を再学習するという方法でかなりの認識率の改善が行なえること、高雑音下ではコードブックマッピングが有効であること、WGD 距離尺度を用いることにより雑音環境下および無雑音環境下における認識率も改善されることを示した。TDNN-LR では、雑音環境下のデータと雑音の無い環境のデータを同時に用いて学習することにより、広範囲のに対して安定した認識率を示すことを確認した。[7]

### 2.3.3 セグメント特徴による耐雑音 HMM

セグメント特徴量を用いた離散分布型 HMM 音声認識 (SQ-HMM) を検討し、音素カテゴリ別に作成するカテゴリ依存作成方法と、認識対象となる音声資料または認識対象に近い発話様式をもつ音声資料から作成する発話様式依存作成方法により作成した SQ コードブック (Segment-Quantization codebook) を用いた発話様式依存 SQ-HMM が、VQ-HMM よりも雑音環境下での音声認識性能が優れていることを示した。[14][44][70]

### 2.3.4 レベル可変雑音モデルによる雑音抑圧法

雑音環境下で安定して動作する音声認識系が望まれており、定常雑音に関する手法、レベル可変の定常雑音に関する手法、非定常雑音に対する手法、堅牢な分析方法が活発に検討されている。本報告では雑音モデルを用いてレベルを推定することにより雑音抑圧を行なう手法について述べる。これは文献 [Mansour-ICASSP1988] の LPC ケプストラムベクトルの線形伸縮雑音モデルを一般化したものである。[26]

### 2.3.5 識別誤り規準を用いた耐雑音 HMM

雑音の問題を扱う場合、雑音を除去する方法や、雑音の特徴を統計的に扱い認識する方法、耐雑音性の高い距離尺度の研究等がなされている。また、このような方法の他に、雑音の特徴を学習することにより、耐雑音性の高いモデルをつくる研究もなされている。一方、識別率を向上させるために識別率最小規準に基づく学習則 (Minimum error classification training (MEC)) が提案され高い認識性能を実現している。しかし、雑音のない環境において MEC 学習されたモデルの耐雑音性の検討はあまりなされていない。また、雑音環境下で学習したモデルは耐雑音性をもつが、学習則 (MEC 学習と最尤推定: Maximum likelihood training (ML)) による比較等もなされていない。本報告では、モデルに連続分布型 HMM を選び、まず ML 学習と MEC 学習との比較により、雑音のない環境において MEC 学習された HMM の雑音重畳資料における認識性能の把握と、雑音の特徴を考慮したパラメータの移動を実現するために、雑音重畳データを用いて MEC 学習を行なうことにより、耐雑音の高い HMM の構築方法を検討する。[27][75]

### 2.3.6 標準 HMM による電話音声認識

標準音声を用いて学習した連続分布 HMM 音声認識システムを用いて、電話音声のように周波数特性の異なる伝送系からの入力音声の認識を行なうための手法を提案し、音素認識実験により性能の改善効果について評価した。大規模な音声データベースが作成されつつあるが、実際の応用においては異なる周波数特性や雑音環境で動作させる必要があるが、そのすべての環境における音声を多量に収集しデータベース化することは一般的には不可能に近い。そこで標準的な特性の音声で作成された音声認識モデルを適応する必要がある。雑音に関して符号帳写像法や VFS 法を提案しその有効性を示した。電話音声の認識についてはその必要性から多くの研究がなされているが、電話音声に

対する適応についてはあまり論じられていない。本手法は電話音声に限らず周波数伝達特性が異なる音声の補正に対して一般的に適用可能であるが電話音声に限定して述べる。[37]

## 2.4 音響特徴量による言語識別の研究

### 2.4.1 20ヶ国語の言語識別

現在音声認識や音声理解、さらに言語間の自動翻訳が活発に研究されている。しかしながら、発話された音声を用いて属する言語を識別する研究は多くはない。言語音の識別に関する研究は比較言語学においてなされているが、言語の基語の決定問題に関連して研究され、おもに文法における類似性、音素系列における類似性が対象であり、音響特徴量からの識別という研究は少ない。また、特に自動認識という観点からの研究はあまり事例は多くはない。本研究では、発話された言語音声の音響特徴量を分析し、言語の自動識別を検討する。[1][56][85]

### 2.4.2 Bilingual 音声を対象とした日英言語識別

現在音声認識や音声理解さらに言語間の自動翻訳が活発に研究されている。しかしながら、発話された音声を用いた言語識別の研究は多くはない。既に筆者は20ヶ国語音声データベースを用いたテキスト独立かつ話者独立言語識別実験を行ないその性能を評価し報告した。本研究では、バイリンガル話者によって発話された日英語の音声を対象とした同一話者内での2言語の識別について検討した。[11][69][89]

### 2.4.3 スペクトル遷移の確率モデルによる言語識別

本研究では同上のタスクに対し、スペクトル遷移の動的な特徴に対してN-gram, ergodic HMMの確率モデルを用いた言語識別の有効性及び雑音環境下での性能の検討を行なった。静的・動的な音声特徴による言語識別法の評価を行なった。動的特徴を用いる方法は静的特徴を用いる方法に比べて、より少ない音声量で識別が可能なこと、また雑音環境下での耐性が強いことがわかった。[20][94]

## 2.5 信号源モデルによる音声の区分化・識別問題の研究

### 2.5.1 話者特徴に基づく音声区分化と識別

従来、話者識別はあらかじめ多量の音声資料を用いて登録(話者モデルの作成)を行ない、その話者モデルを用いて認識を行なうものである。本研究では、音声に含まれる複数話者の発話を事前登録なしに識別する問題を検討した。入力系列中での話者の遷移位置(話者 segmentation 問題)、および話者数 $N$ は既に得られていると仮定し、Universal VQ 符号の出力確率分布の $N$ 個のカテゴリへのクラスタリングによる方法を提案した。セグメンテーションが未知の場合についてはHMMの適用した。複数話者の発話を事前登録なしに識別する問題を検討し、その有効性を示した。複数話者による対話音声の話者の遷移問題などに応用が期待される。[25][49][50][73][80][38][97]

### 2.5.2 男女性差に基づく音声区分化と識別

話者特徴に基づく発話の区分化・分類問題に対し解法が提案され、実験を通して有効性が評価されている。本研究は、同一の枠組を用いた男女性の音声特徴に基づく音声セグメンテーションとクラスタリングに関する検討である。音声による男女性識別の研究はあまり多くはない。さらに、従来の男女性話者識別問題の研究はあらかじめ登録したパターンや学習したガウス識別器などに基づく”登録型”の男女性の識別問題のみであり、与えられた音声のみを用いた男女性の識別は皆無である。一方、男女声識別問題は2つの信号源への区分化・分類問題であり、筆者等が検討してきた、一般の $N$ 個の信号源への区分化・分類問題の縮退問題と見ることができ。即ち、 $N$ 話者からの問題の場合、信号源の数 $N$ を縮退させ話者信号源の縮退信号源が男女に対応するかという問題を検討することになる。出力確率ベクトルクラスタリング法の評価を行ない有効性を明らかにした。男女話者が最大8名の小規模な実験ではあるが、話者の数に依らず少数単語で良好に識別できることが分かった。[33]

## 2.6 その他

### 2.6.1 歪み尺度測地線を用いた音声スペクトルの補間

音声をARモデルのパラメータを用いてユークリッド空間の点と見ることができ。スペクトル距離尺度(歪み尺度)がユークリッドであれば直線が2点間の最短経路を与えるが、そうでなければ直線とは限らない。歪み尺度は

空間に歪みをもたらし、最短経路も歪んでしまう。以下では歪み尺度による最短経路を'歪み尺度測地線'と呼び、その性質を述べ音声への応用を述べた。[2][42] [55][76][91]

### 2.6.2 伝送誤りに強い VQ 符号帳構成法

ベクトル量子化 (VQ) は音声符号化・音声認識において用いられている強力な情報圧縮手法である。伝送誤りの伴う音声伝送系においては VQ 符号のビット誤りは復号時に歪みを発生する。この歪みをできるだけ小さくするように VQ 符号帳を再設計することを検討した。即ち、ビット表現された VQ 符号に対して確率的に発生するビット誤りに伴う歪みを最小化するように VQ 符号のビット表現を決定する。これは確率と歪みとを同時に考慮した確率歪み量を目的関数とする組合せ問題である。組合せが有限であるので有限回の探索で最適解が求められるが、符号帳が現実に用いられている程度の大きさの場合その可能な組合せが膨大になるため最適解の探索は容易ではない。ここではその最適解の探索に Hopfield Model などの Neural Network Model を適用することとし、その基礎的な検討を行なった。[93]

### 2.6.3 LPC ケプストラム係数の存在領域の諸性質

本研究では、LPC ケプストラム係数ベクトルの存在領域に関する幾つかの性質について述べた。LPC ケプストラム係数は、対数 LPC スペクトルのフーリエ係数と言う物理的な意味を持ち、聴覚特性にも合致したパラメータであると同時に、その計算の簡単さから音声認識・符号化等の分野で広く用いられている。本研究では LPC ケプストラム係数ベクトルの存在領域に関する有界性、連結性、対称性、包含関係、分析次数に関する増加性、加法分解性、非凸性について述べた。さらに LPC ケプストラムベクトルの時間変化の特徴量である  $\Delta\text{CEP}$  の性質についても述べた。[6][46] [79]

### 3 発表論文題目一覧

以下に発表論文、特許などの題目一覧を示す。

- 日本音響学会講演論文
- 電子情報通信学会講演論文
- 音声研究会発表論文
- 国際会議発表論文
- 学術論文
- ATR テクニカルレポート
- 講演
- 特許
- ATR Working Group

#### 日本音響学会講演論文

#### 参考文献

- [1] 杉山, 音響特徴量を用いた多言語音声識別の検討, 音響学会講演論文集, 3-3-6, pp.81-82 (1990-03).
- [2] 杉山, 歪み尺度測地線を用いた音声の幾何的な構造について, 音響学会講演論文集, 3-4-16, pp.299-300 (1990-03).
- [3] 杉山, 福沢, 沢井, 嵯峨山, ニューラルネットによる集合間写像の教師なし学習, 音響学会講演論文集, 2-P-10, pp.149-150 (1990-09).
- [4] 大倉, 杉山, 波形入出力による雑音抑圧ニューラルネットの音声認識への応用, 音響学会講演論文集, 1-8-3, pp.5-6 (1990-09).
- [5] 福沢, 沢井, 杉山, ニューラルネットワークによる恒等写像を用いた話者適応, 音響学会講演論文集, 1-8-16, pp.31-32 (1990-09).
- [6] 杉山, LPC ケプストラム係数の存在領域について, 音響学会講演論文集, 2-6-4, pp.253-254 (1991-03).
- [7] 大倉, 杉山, 雑音環境下における HMM と TDNN の文節認識性能の評価, 音響学会講演論文集, 2-5-9, pp.69-70 (1991-03).
- [8] 藤原, 小森, 杉山, HMM とスペクトログラムリーディング知識に基づくハイブリッド音素セグメンテーションシステムの構想, 音響学会講演論文集, 2-5-16, pp.85-86 (1991-03).
- [9] 沢井, 中村, 福沢, 杉山, ニューラルネットワークによる不特定話者音声認識へのアプローチ法について, 音響学会講演論文集, 1-5-17, pp.37-38 (1991-03).
- [10] 杉山, Alain Biem, TDNN-HMM による音素認識, 音響学会講演論文集, 2-P-6, pp.149-150 (1991-10).
- [11] 杉山, Bilingual 音声を対象とした日英言語識別の検討, 音響学会講演論文集, 3-6-12, pp.133-134 (1991-10).
- [12] 福沢, 小森, 沢井, 杉山, セグメントベース話者適応ニューラルネットワークを用いた文節音声認識, 音響学会講演論文集, 3-5-10, pp.109-110 (1991-10).
- [13] 福沢, 杉山, ニューラルネットワークによる教師なし話者適応法とその評価, 音響学会講演論文集, 3-5-11, pp.111-112 (1991-10).

- [14] 大倉, 杉山, セグメント特徴量を用いた雑音環境下での HMM 音声認識の検討, 音響学会講演論文集, 1-5-6, pp.11-12 (1991-10).
- [15] 藤原, 岩橋, 小森, 杉山, HMM とスペクトログラムリーディング知識に基づくハイブリッド音素セグメンテーションシステム, 音響学会講演論文集, 2-5-20, pp.87-88 (1991-10).
- [16] 中村, 沢井, 杉山, TDNN を用いた不特定話者音素認識, 音響学会講演論文集, 2-5-8, pp.63-64 (1991-10).
- [17] 小森, 福沢, 杉山, A.H.Waibel, 嵯峨山, ニューラルファジー学習の連続音声認識における効果, 音響学会講演論文集, 2-5-11, pp.69-70 (1991-10).
- [18] 岩橋, 藤原, 小森, 杉山, 匂坂, 自動セグメンテーションによる音声合成単位の作成, 音響学会講演論文集, 1-6-21, pp.231-232 (1991-10).
- [19] 杉山, 栗並, 識別誤り最小規準による話者写像ニューラルネットの最適化, 音響学会講演論文集, 2-Q-22, pp.201-202 (1992-03).
- [20] 杉山, 吉田, スペクトル遷移の確率モデルによる言語識別の検討, 音響学会講演論文集, 1-P-1, pp.105-106 (1992-03).
- [21] 福沢, 杉山, TDNN-LR 連続音声認識における不特定話者 TDNN と話者適応ニューラルネットワークの性能評価, 音響学会講演論文集, 2-Q-21, pp.199-200 (1992-03).
- [22] 加藤, 杉山, 多入力素子をもつニューラルネットワークを用いた連続音声認識, 音響学会講演論文集, 3-1-1, pp.71-72 (1992-03).
- [23] 大倉, 土井, 杉山, 移動ベクトル場平滑化話者適応方式を用いた雑音環境下音声認識, 音響学会講演論文集, 2-Q-16, pp.189-190 (1992-03).
- [24] 大倉, 杉山, 嵯峨山, 混合連続分布 HMM を用いた移動ベクトル場平滑化話者適応方式, 音響学会講演論文集, 2-Q-17, pp.191-192 (1992-03).
- [25] 杉山, 村上, 渡辺, 話者特徴に基づく音声セグメンテーションとクラスタリング, 音響学会講演論文集, 2-Q-21, pp.213-214 (1992-10).
- [26] 杉山, 大倉, レベル可変雑音モデルによる雑音抑圧法, 音響学会講演論文集, 1-7-12, pp.69-70 (1992-10).
- [27] 大倉, D.Rainton, 杉山, 識別誤り規準を用いた耐雑音 HMM の検討, 音響学会講演論文集, 1-7-14, pp.73-74 (1992-10).
- [28] 大倉, 杉山, 嵯峨山, 混合連続分布 HMM 移動ベクトル場平滑化話者適応方式の文節認識性能の評価, 音響学会講演論文集, 2-5-4, pp.125-126 (1992-10).
- [29] 福沢, 加藤, 杉山, FPM-LR を用いた不特定話者連続音声認識, 音響学会講演論文集, 3-7-3, pp.167-168 (1992-10).
- [30] 加藤, 杉山, ニューラルネットワークを用いた不特定話者の特徴抽出について, 音響学会講演論文集, 2-Q-20, pp.211-212 (1992-10).
- [31] 藤原, 杉山, 音素ラベリングワークベンチ, 音響学会講演論文集, 2-Q-28, pp.227-228 (1992-10).
- [32] 山口, 他, ATREUS:ATR における連続音声認識方式の比較, 音響学会講演論文集, 2-Q-5, pp.181-182 (1992-10).
- [33] 杉山, 村上, 男女性音声特徴に基づく音声セグメンテーションとクラスタリング, 音響学会講演論文集, 1-4-9, pp.17-18 (1993-3).
- [34] 福沢, 杉山, ニューラルネットワークを利用した予備選択による大語彙単語音声認識, 音響学会講演論文集, 2-Q-20, pp.123-124 (1993-3).

- [35] 藤原, 杉山, HMM とスペクトログラムリーディング知識に基づく不特定話者音素セグメンテーションの連続音声における評価, 音響学会講演論文集, 2-Q-12, pp.107-108 (1993-3).
- [36] 大倉, 杉山, 嵯峨山, 連続分布 HMM 移動ベクトル場話者適応方式の 1 検討, 音響学会講演論文集, 2-Q-17, pp.117-118 (1993-3).
- [37] 片岸, 藤原, 杉山, 標準音声学習連続分布 HMM の電話音声認識への適用, 音響学会講演論文集, 2-Q-15, pp.113-114 (1993-3).

#### 電子情報通信学会講演論文

- [38] 杉山, 渡辺, 村上, 事前登録なしの話者識別問題, 電子通信学会部門別全国大会, SA-6.1, 1-263-264 (1992-09).
- [39] 福沢, 杉山, 階層的クラスタリングと Neural Network を用いた教師なし話者適応法, 電子通信学会部門別全国大会, SA-6.5, 1-271-272 (1992-09).
- [40] 加藤, 杉山, 多入出力素子をもつニューラルネットワークを用いた連続音声認識, 電子情報通信学会「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」時限研究専門委員会, SPREC-91-2, pp.47-48 (1992-02).
- [41] 大倉, 杉山, 嵯峨山, 混合連続分布 HMM を用いた移動ベクトル場平滑化話者適応方式, 電子情報通信学会「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」時限研究専門委員会, SPREC-91-2, pp.49-50 (1992-02).

#### 音声研究会発表論文

- [42] 杉山, 歪み尺度測地線を用いた音声スペクトルの補間, 音声研究会資料, SP90-03, pp.17-24 (1990-05).
- [43] 大倉, 服部, 杉山, 鹿野, コードブックマッピングを用いた雑音下での音声認識, SP90-12, pp.25-32 (1990-06).
- [44] 大倉, 杉山, セグメント特徴量を用いた雑音環境下での HMM 音声認識, SP91-55 (1991-09).
- [45] 福沢, 小森, 沢井, 杉山, セグメントベース話者適応ニューラルネットワークと TDNN-LR を用いた文節音声認識, SP91-105, pp.23-30 (1992-1).
- [46] 杉山, LPC ケプストラム係数の存在領域について, SP91-120, pp.17-24 (1992-02).
- [47] 加藤, 杉山, ファジーパーティションモデルを用いた連続音声認識, SP92-28, pp.31-38 (1992-06).
- [48] 大倉, 杉山, 嵯峨山, 混合連続分布 HMM を用いた移動ベクトル場平滑化話者適応方式, SP92-16, pp.23-28 (1992-06).
- [49] 渡辺, 村上, 杉山, 未知・複数信号源クラスタリング問題 — 未知話者クラスタリングへの応用 —, SP92-45, pp.47-54 (1992-07).
- [50] 村上, 杉山, 渡辺, HMM を用いた未知・複数信号クラスタリング問題の検討, SP92-74, pp.7-14 (1992-10).
- [51] 福沢, 加藤, 杉山, FPM-LR を用いた不特定話者連続音声認識の実現, SP92-107, pp.31-38 (1992-12).
- [52] I.Donescu, 加藤, 杉山, ニューラルネットワークを用いた不特定話者の特徴抽出と音声認識によるその評価, SP92-117, pp.15-22 (1993-1).
- [53] 永井, 他, ATR における連続音声認識システム "ATREUS" の諸方式と性能, SP92-122, pp.51-58 (1993-1).

- [54] 藤原, 杉山, HMM とエキスパートシステムの手法を用いた音素ラベリングワークベンチ, SP92-132, pp.73-80 (1993-1).

国際会議発表論文

- [55] M.Sugiyama, Spectral Interpolation Using Distortion Geodesic Lines, ICLSP90, 11.15, pp.477-480 (1990-11).
- [56] M.Sugiyama, Automatic Language Recognition Using Acoustic Features, ICASSP91, 16.S12.8, pp.813-816 (1991-05). (at Tronto, Canada)
- [57] K.Ohkura, M.Sugiyama, Speech Recognition in a Noisy Environment Using a Noise Reduction Neural Network and a Codebook Mapping Technique, ICASSP91 (1991-05).
- [58] M.Sugiyama, H.Sawai, A.H.Waibel, Review of TDNN (TIME DELAY NEURAL NETWORK) Architectures for Speech Recognition, Proc. of ISCAS91 (Jun. 1991) (Singapore).
- [59] K.Fukuzawa, H.Sawai, M.Sugiyama, Segment-Based Speaker Adaptation by Neural Network, Proc. of Neural Networks for Signal Processings, pp.440-451 (1991-10).
- [60] Fujiwara, Y.Komori, M.Sugiyama, A Hybrid System by Phoneme Labelling Based HMM and Spectrum Reading Knowledge,
- [61] S.Nakamura, H.Sawai, M.Sugiyama, Speaker-Independent Phoneme Recognition Using Large-Scale Neural Networks, Proc. of ICASSP92, 46.5, pp.I-409-412 (March 1992).
- [62] K.Fukuzawa, Y.Komori, H.Sawai, M.Sugiyama, A Segment-Based Speaker Adaptation Neural Network Applied to Continuous Speech Recognition, Proc. ICASSP92, 55.1, pp.I-433-436 (March 1992).
- [63] S.Fujiwara, Y.Komori, M.Sugiyama, An Integrated System for Automatic Labelling Based on HMM and Spectrogram Reading Knowledge, ISSPA92 (1992-08).
- [64] Y.Kato, M.Sugiyama, Fuzzy Partition Models and Their Effect in Continuous Speech Recognition, NNSP92, pp.111-120 (1992-8).
- [65] M.Sugiyama, K.Kurinami, Minimal Classification Error Optimization for a Speaker Mapping Neural Network, NNSP92 Workshop, pp.233-242 (1992-8).
- [66] K.Fukuzawa, Y.Kato, M.Sugiyama, A Fuzzy Patition Model(FPM) Neural Network Architecture for Speaker Independent Continuous Speech Recognition, ICSLP92, Th.PM.P.14, pp.1383-1386 (1992-10).
- [67] K.Ohkura, M.Sugiyama, Shigeki Sagayama, Speaker Adaptation Based on Vector Field Smoothing with Continuous Mixture Density HMMs, ICSLP92, We.fAM.1.1, pp.369-372 (1992-10).
- [68] S.Fujiwara, Y.Komori, M.Sugiyama, A Phoneme Labelling Workbench using HMM and Spectrogram Reading Knowledge, ICSLP92, Fr.fAM.2.2, pp.791-794 (1992-10).
- [69] M.Sugiyama, Language Recognition Using Spectral Features, Speech Research Symposium, Rutgers University, 採録決定 (1992-06).
- [70] K.Ohkura, M.Sugiyama, Noisy Speech Recognition with Codebook Mapping and Segmental Approaches, IEEE Workshop on Interactive Voice Technology for Telecommunications Applications, 採録決定 (1992-10).
- [71] A.Biem, M.Sugiyama, A Hybrid Stochastic Connectionist Approach to Automatic Speech Recognition, Proc. of 1st African Conference on Research in Computer Science, Vol.2, pp.605-617 (Oct. 1992).

- [72] S.Sagayama, M.Sugiyama, et al, ATREUS: Continuous Recognition Systems at ATR Interpreting Telephony Research Laboratories, SST92 (1992), 採録決定.
- [73] M.Sugiyama, J.Murakami, H.Watanabe, Speech Segmentation and Clustering Based on Speaker Features, ICASSP93 (1993) 論文掲載予定.
- [74] Y.Katoh, M.Sugiyama, Speaker-Independent Features Extracted by A Neural Network, ICASSP93 (1993) 論文掲載予定.
- [75] K.Ohkura, D.Rainton, M.Sugiyama, Noise Robust HMMs Based on Minimum Error Classification Training, ICASSP93 (1993) 論文掲載予定.

### 学術論文

- [76] M.Sugiyama, Distortion Geodesic Lines and Their Application to Spectral Interpolation, IEICE Trans., Vol.E-74, No.3, pp.609-614 (Mar. 1991).
- [77] Y.Kato, M.Sugiyama, Fuzzy Partition Models and Their Incremental Training for Continuous Speech Recognition, ASJ(E), Vol.13, No.6, pp.411-418 (1992 Nov.).
- [78] K.Kurinami, M.Sugiyama, An Optimization Technique for Speaker Mapping Neural Networks Using Minimal Classification Error Criterion, ASJ(E), Vol.13, No.6, pp.419-428 (1992 Nov.).
- [79] 杉山 雅英, LPC ケプストラムの存在領域の諸性質, 電子情報通信学会論文誌 A 論文投稿中.
- [80] 杉山 雅英, 村上 仁一, 渡辺 秀行,  $N$  信号源モデルに基づく音声の区分化識別問題, 電子情報通信学会論文誌, DII 論文投稿中.
- [81] 福沢 圭二, 加藤 喜永, 杉山 雅英, FPM-LR による不特定話者連続音声認識, 電子情報通信学会論文誌, DII 論文投稿中.
- [82] 大倉, 杉山, コードブックマッピング手法を用いた雑音環境下音声認識, 電子情報通信学会論文誌, DII 論文投稿予定.
- [83] 大倉, 嵯峨山, 杉山, 混合連続分布 HMM 移動ベクトル場平滑化話者適応方式, 電子情報通信学会論文誌, DII 論文投稿中.
- [84] 大倉, D.Rainton, 杉山, 電子情報通信学会論文誌, DII 論文投稿予定.

### ATR テクニカルレポート

- [85] 杉山, Automatic Language Recognition Using Acoustic Features, TR-I-0167.
- [86] 杉山, ATR における Neural Network を用いた音声情報処理, TR-I-0173.
- [87] A.Biem, M.Sugiyama, Study on Combining HMMs and Neural Network Models, TR-I-0174 (1991-02).
- [88] S.YAMAFUKU, M.SUGIYAMA, TDNN-HMM-LR による文節音声認識, TR-I-0211 (1991-04).
- [89] H.FUJIWARA, M.SUGIYAMA, バイリンガル音声を用いた日・英語の言語識別, TR-I-0212 (1991-04).
- [90] 中村, 沢井, 杉山, TDNN を用いた不特定話者の音素認識 (2), TR-I-0219 (1991-06).
- [91] 杉山, 歪み尺度測地線を用いた音声スペクトルの補間, TR-I-0221.
- [92] 山本, 福沢, 杉山, VQ Neural Network による教師なし話者適応, TR-I-0222. (1991-07).
- [93] 杉山, 嵯峨山, 伝送誤りに強い VQ 符号帳構成法に関する検討, TR-I-0223 (1991-08).

- [94] 吉田, 杉山, 確率モデルを用いた言語識別の検討, TR-I-0226 (1991-09).
- [95] 栗並, 杉山, 話者写像ニューラルネットの識別誤り最小基準による最適化, TR-I-0227 (1991-09).
- [96] 安岐, 加藤, 杉山, Radial Basis Function を適用した Fuzzy Partition Model による音声認識, TR-I-0249 (1992).
- [97] 渡辺, 村上, 杉山,  $N$  信号源分解問題, TR-I-0262 (1992-04).
- [98] 大倉, 杉山, 話者選択手法を用いた音声認識の基礎検討, A study of speech recognition using speaker selection approach, TR-I-0265 (1992-).
- [99] 大倉, 杉山, 嵯峨山, 混合連続分布HMMを用いた移動ベクトル場平滑化話者適応方式, Speaker adaptation based on transfer vector field smoothing model with continuous mixture density HMMs, TR-I-0266 (1992-).
- [100] Franck Martin, M.Sugiyama, S.Sagayama, Composition of Noise and Clean-Speech HMMs for Recognition of Noisy Speech, TR-I-0283 (1992-12).
- [101] Ioana Donescu, 加藤 喜永, 杉山 雅英, Speaker-Independent Features Extracted by a Neural Network and Their Evaluation in Speech Recognition, TR-I-0296 (1993-02).
- [102] 杉山, 音声情報処理研究室 デモソフトウェア 解説書, TR-I-0299 (1993-02).
- [103] 杉山, 音声情報処理研究室 計算機・ソフトウェア 解説書, TR-I-0300 (1993-02).
- [104] 藤原 紳吾, 杉山 雅英, HMMとエキスパートシステムの手法を用いた音素セグメンテーションシステムとラベリングワークベンチ, TR-I-0305 (1993-03).
- [105] 片岸, 藤原, 杉山, マイク入力音声で学習した混合連続分布 HMM の電話音声認識への適用, TR-I-0306 (1993-03).
- [106] 杉山, 3年間の音声情報処理研究をふりかえって -音声情報処理研究および研究支援業務の概観-, TR-I-0309 (1993-03).
- [107] 藤原, 杉山, Labelling Workbench -User's Guide -, TR-I-0309 (1993-03).

### ATR Working Group

- [1] 杉山, 識別誤り最少規準による話者適応写像ニューラルネットの学習, ATR Working Group (1992-01).

### ATR テクノロジーシリーズ

- [1] “自動翻訳電話”, ATR テクノロジーシリーズ, No.1.
- [2] “ニューラルネットワーク応用”, ATR テクノロジーシリーズ, No.3.

### 講演

- [1] 杉山, Neural Network を用いた音声情報処理, 神経情報科学特選講座, 第9回「ニューロモデルの新しい流れ」(1990-09). 財団法人千里国際情報事業財団.

- [1] 大倉, 杉山, 雑音除去と話者適応の機能を有する音声認識装置, (1991-09). 公報公開済
- [2] 福沢, 沢井, 杉山, ニューラルネットワークによる話者適応化方式, (1990-10).
- [3] 杉山, 音声中の複数話者の発話区間自動検出同定方式, (1992-8)
- [4] 福沢, 杉山, ニューラルネットワークの発火パターンを用いた単語予備選択方式, (1993-02).

以上の発表論文などの件数を以下に示す。

	種別	件数	記事
講演発表	日本音響学会	37	筆頭 11, 連名 26
	電子情報通信学会	4	筆頭 1, 連名 3
	音声研究会	13	筆頭 2, 連名 11
	小計	54	
論文	国際会議	21	筆頭 6, 連名 15
	学術論文	9	筆頭 3, 連名 6
	小計	30	
ATR	Technical Report	23	筆頭 7, 連名 16
	Working Group	1	
	先端テクノロジーシリーズ分担執筆	2	
	小計	26	
	総計	110	
その他	講演	2	
	特許	4	

## 4 研究支援業務概観

以下に研究項目以外の研究室運営の項目に関してまとめ、今後の方針などを述べる。

### 4.1 研究室運営業務

#### 4.1.1 研究室 Regular Talk の運営

研究室の研究進捗に合わせて定期的に Regular Talk (SP) を開催した。原則的に木曜日の午後1時10分から2時間位に行なった。質問時間が長くないように司会・運営が望まれる。これは視聴覚機構研究所 片桐氏と共同運営であるが実質的には当研究室が中心となっている。音響学会などへの研究員の発表のリハーサルについてもこの枠の中で計画の立案・実行管理を行なった。これらの日程の周知はメール及びニュースで行なった。今後はニュースに一元化することが望まれる。また、計画の立案は良いとして周知作業などは研究室秘書に依頼したい。

#### 4.1.2 Final Talk, 実習生報告会の運営

出向元に復帰する研究員の Final Talk (分かれのご挨拶) 及び実習生報告会については社長の日程を確認し、21,22 会議室にて開催した。鷹見研究員に写真の撮影を依頼した。会議室のレイアウトの変更(机、椅子の並べ換え)については研究員で分担して作業した。

#### 4.1.3 各種研究室日程の作成と室員への周知

学会への proposal の提出、学会発表のリハーサルなどの種々の日程を調整し、研究室員へ周知徹底した。以下にその例を示す。

例：7 - 10月の予定

項目	締切	研究室予定
SST92 proposal	7/27	7/21(Tue)15:00
ICASSP93 proposal	8/3	7/31(Fri)15:00
音響学会原稿締め切り	8/5(Wd)	8/3(Mon)
ISSPA92 rehearsal	8/16-21	8/7(Fri)
NNSP92 rehearsal	8/31-9/2	8/24(Mon)
9月音声研究会原稿	8/9?	
9月音声研究会 rehearsal		9/4(Fri)
9月音声研究会	9/9-11	9/9-11
通信学会 rehearsal	9/24	
音響学会 rehearsal	9/21-23	
通信学会大会	9/27-30	9/27-30
音響学会大会	10/5-7	10/5-7
ICSLP92 rehearsal		10/1,2
ICSLP92	10/12-16	10/12-16
10月音声研究会	10/22,23	10/22,23
ATR 研究発表会	11/10	

#### 4.1.4 アルバイト学生の確保

夏休みなどを利用した短期のアルバイトの学生を用いて、研究の補助作業や研究環境の整備を行なった。京都工芸繊維大学の小林豊先生にお願いして、学内の掲示板にアルバイト募集の広告を出していただいたり、研究室の学生を紹介していただいたりして確保した。アルバイト期間中に数度のヒアリングを行ない、問題点の整理を行なうようにした。

#### 4.1.5 実習生の確保

実習生が少ない場合に音声認識を研究している幾つかの大学の先生方に手紙を出し、実習生の紹介を要請した。慶応大学、早稲田大学、竜谷大学、豊橋技術科学大学、INT、ENST などの大学からは定期的に学生を送り出していただくことができた。これらの協力関係を今後も維持していく必要がある。

#### 4.1.6 ATR Working Group の主催

ATR Working Group の開催を分担して行なった。筆者は第2種研究会“連続音声認識シンポジウム”の主催を担当した。

#### 4.2 研究所運營業務

以下の委員会などでの運營業務を分担した。

- プロジェクト終了委員会
- 研究発表会 デモ作成管理
- CSTAR 実行委員会 報道担当
- 次期プロジェクト(新自動)の計画立案

#### 4.3 研究室 計算機及びソフトウェアの整備

##### 4.3.1 計算機環境の整備

計算機環境の整備のため計算機委員会を招集し計算機環境の整備・問題点の解決を行なった。これは各 VAX 系計算機などのシステム管理者及び興味のある研究員で構成した。今後もこのような委員会が必要であろう。

委員：嵯峨山, 杉山, 福沢(sp), 大倉(la), 鷹見(ap), 三村(fs), 藤原(as04), 海木, 小坂(hp), 加藤(mac), 小野(DEC)

研究員の移動の時期には計算機(ワークステーション)や研究員の机の場所などについて不足の生じないように準備を行なった。作成には担当する研究員と事前に相談し、“最大多数の最大幸福の原理”に従った。作成例を以下に添付する。

例：一時研究員座席表

Name	Period	Tutor	Office
Ioanna(INT)	8/1 - 1/31	Kato	Labeling office (Zhou)
?(ENST)	8/1 - 1/31	Kosaka	Wan's office
Wan (TL)	- 8/4	Sagisaka	using
Jaq	7/13 - 9/13	Sagisaka	old Tagawa's office
Mizuno(Waseda Univ.)	8/1 - ?	Singer	Singer's office
Paul	- 11/8	Sagisaka	using
Yamamoto(Waseda)	- ?	Murakami	using
Martin	8/31 - 10/2	Ohkura	Jaq's office
Awatsu(Tohoku Univ)	8/3 - 9/6	Isotani	Labeling office (Kinugasa)

##### 4.3.2 DEC 定例打ち合せ

DEC 計算機に関する定期的(1月に一回)打ち合せを行なった。この打ち合せで DEC 系計算機・ネットワークに関する問題点の検討、研究員からの要望・苦情の解決、今後の DEC 計算機販売計画などの情報入手、などを行なった。

担当：山口 毅(日本 DEC 常駐 SE), 三浦(DEC)

##### 4.3.3 デモプログラムの作成支援

伴, 高嶋, 田川(SET) 君により音声情報処理研究室の最新の研究結果を計算機上でデモが行なうようにした。デモの本体は研究者が開発し、主に表示部分(X11 or motif 使用)の作成を担当した。作成には従来 DECwidget を用いていたが、新たに作成するものは可搬性を重視して Xlib のみを用いることとし、DECwidget で書かれている物については motif による書換えをおこなった。また彼らプログラマーにはプロジェクト終了前に行なわれた CSTAR デモにおけるソフトウェアの開発にも大きく貢献をしてもらった。

デモの作成順番などは杉山が窓口となり管理を行なった。作成順番の混乱を避けるため、研究者からの直接の作成依頼は原則行なわせないように指導した。今後もこのデモ開発体制は必要であろう。また、窓口となる管理者が必要であると思われる。

担当者：伴, 高嶋, 田川 博章 (SET)

#### 4.3.4 外部ソフトハウスによる計算機ソフトの作成

NOVA に依頼してソフトウェアの開発、修正を行なった。

##### 1. Recurrent Network のプログラム

DEC Workstation (5000 かつ 3100) を含む他の計算機で汎用的に動作可能なリカレント型ニューラルネットワークの学習及び評価用プログラムの作成を行なった。この際、従来から ATR で開発してきたニューラルネットワーク関連プログラムとの整合性を図るものとした。

##### 2. Recurrent Network のプログラム (2)

既に作成したリカレント型ニューラルネットワークの学習及び評価用プログラム、さらにこれを制御するニューラルネットワーク実験用のソフトウェア dcp2 において、ATR が指示したアルゴリズムに基づき高速化を行なう。この高速化処理は DEC Workstation (5000 かつ 3100) を含む他の計算機で汎用的に動作可能であるものとする。

##### 3. ニューラルネット用汎用グラフィックプログラム

既に作成したニューラルネットワークの設計・学習及び評価用のプログラム (nnwb4) において使用されている DEC X window tools (widget) を、より汎用的な motif もしくは X11 に書き換え、DEC 以外のワークステーションにおいて動作可能ならしめる。この場合、従来のプログラムの動作・処理速度などは同一性を保つものとする。

#### 4.3.5 計算機ファイルバックアップ業務

計算機ディスクの定期的なバックアップを行なった。主に VAX 系計算機、ファイルサーバー及び一部のデモ専用機のディスクを対象とした。当初、MT を用いていたが、最終的に当研究室は DAT による file system 指定のバックアップに切替え、作業の効率化を図った。今後はより効率良いバックアップ体制を検討する必要がある。

担当：橋直子, 中井 隆浩 (ユーニス株式会社)

#### 4.3.6 デモプログラムの解説書作成

作成デモに関する説明資料の作成を行なった。hostmachine, directory, file の所在、make の仕方、run の仕方、関連する研究発表文献などを述べた。今後もデモを作成する毎にこのような簡単な解説を作成することが望ましいように思われる。詳しくは文献 [102] を参照すると良い。

担当者：田川 博章 (SET)

#### 4.3.7 ソフトウェア解説書作成

PDS や研究者作成の各種のソフトウェアにより研究室の計算機ソフトウェア環境は随分整備された。この中で特に藤原, 海木研究員の功績は大きい。ソフトウェアに関しては文献 [103] を参照すると良い。

担当者：山口 毅 (日本 DEC), 加藤初子 (UNISS)

#### 4.3.8 T<sub>E</sub>Xstyle file の整備

T<sub>E</sub>X の普及の促進するために各種のスタイルファイルの整備を行なった。スタイルファイルの種類などについては文献 [103] の p.39 に述べてある。これ以外に電子情報通信学会論文投稿用スタイル IEICE.sty などがある。今後も必要に従って体系的に整備する必要がある。

#### 4.3.9 文献データベースの作成整備

関連の研究論文文献のデータベースの整備を行なった。これは研究室の財産ともなるものであるため、組織的に体系的に行なう必要がある。また、整備の形式についても個人の趣味に走ることなく汎用性のある BiB<sub>T</sub>E<sub>X</sub> 形式などで整備する必要がある。一部整備した部分については文献 [103] の p.75 に述べた。

## 4.4 講演会、分担執筆及びシンポジウムの運営

### 4.4.1 技術講演の講師

#### 1. “ニューラルネットワークによる音声情報処理”

Neural Network Model が音声情報処理に応用され、種々の分野においてその有効性が報告されている。本資料では Neural Network Model を用いた音声情報処理の研究を概観する。本報告では主に 1989 年及び 1990 年に発表された論文について概観している。言語処理への応用、エキスパートシステムとの融合、理論的な解析に関する研究成果については割愛した。[1]

#### 2. “ニューラルネットワークの音声認識への応用”

Neural Network Model が音声情報処理に応用され、種々の分野においてその有効性が報告されている。本講演ではニューラルネットワークを用いた連続音声認識の研究の現状について述べ、TDNN(Time-Delay Neural Network) と LR parser と用いた連続認識システムについて、その構成方法、学習方法、音素認識、文節音声認識性能について述べる。さらに、特定話者 TDNN-LR 認識系を不特定話者音声認識へと発展させる 2 つの方法についての現状、ネットワークの構造の面からと教師信号の面からの TDNN の改良方法とその性能、教師信号の自動生成に基づく学習法、識別誤り最小基準による新たな学習法、高性能な認識系の構築について述べた。[2]

### 4.4.2 ATR テクノロジーシリーズの執筆分担

#### 1. 自動翻訳電話

“ニューラルネットワークによる日本語音素認識”、“ニューラルネットワークによる連続音声認識”、“ニューラルネットワークを用いた話者適応”を執筆分担した。

#### 2. ニューラルネットワークの応用

“TDNN の大語彙単語音声認識への適用”、“不特定話者音声認識ニューラルネットワーク”、“話者適応ニューラルネットワーク”、“Fuzzy Partition Model(FPM)”、“言語処理への応用”、“音声特徴抽出への応用”を執筆分担した。

### 4.4.3 第 2 種研究会 “連続音声認識シンポジウム” の運営

連続音声の認識を目的とした音声処理と言語処理、およびそれらのインタラクションについての最近の研究展開に関する一般公募発表のセッションからなる。公募発表は音声・言語処理のインタラクションに関する研究成果、問題提起、サーベイなど。研究室内で担当委員会を作成して分担して作業を進めた。

- 日時：1992年2月7日(金) - 8日(土)
- 会場：ビューホテル開華亭(静岡県浜松市館山寺町 館山寺温泉)

## 4.5 音声研究会幹事

音響学会、電子情報通信学会音声研究会幹事として、音声研究会の開催および以下のワークショップなどを他の幹事と分担して行なった。

### 4.5.1 ATR での音声研究会の主催

研究室内に委員会を作成して以下の項目について分担して作業を進めた。

会場 確保(杉山)、学会への連絡事項作成(杉山)、ATR バスの便の確認(柳原)、公共バスの便の確認(柳原)、食堂の確認(福沢)、懇親会(福沢)、オープンハウスの準備(嵯峨山、藤原、加藤)、会計係(福沢、柳原)

- 日時  
1月18日(月) 9:30-17:10  
1月19日(火) 9:00-16:10
- 会場 ATR 2階 21, 22 会議室

#### 4.5.2 “ビギナーのための音声処理ワークショップ”の企画

研究の第一線で御活躍の講師による基礎技術のワークショップ（無料）を開催した。この分野の現状を知りたい方、対話音声の認識・理解の研究をこれから始めたい方、初心者に限らず経験豊富な研究者の方など、多くの参加者を対象とした。開催状況の表1を以下に示す。

表 1: 音声処理ワークショップの開催

6 月研究会 「対話音声の認識・理解」			
ATIS システム	古井	NTT	furui@speech-convex.NTT.jp
対話システム	伊藤	東京工業大学	itou@cs.titech.ac.jp
マルチモーダル	竹林	東芝	yoichi@isl.rdc.toshiba.co.jp
A* ALGORITHM	河原	京都大学	kawahara@kuis.kyoto-u.ac.jp

  

9 月研究会 「音声の生成・合成・認識」			
音声生成	鍋木	NTT	kabu@voice-sun.NTT.jp
音声合成	野村	NTT	nomura@speech-convex.NTT.jp
音声認識	鷹見	ATR	jun@atr-la.atr.co.jp

  

10 月研究会 「信号処理・聴覚生理・ニューラルネットワーク」			
音響信号処理	王	名古屋	wang@itakura.nuee.nagoya-u.ac.jp
聴覚	平原	NTT	hirahara@av-sun2.ntt.jp
Wavelet	入野	NTT	irino@nttlab.ntt.jp
Recurrent Neural Network	船橋	豊橋	なし

#### 4.5.3 “国際音声処理ワークショップ (IWSP93)”の企画

ワークショップは音声研究会が主催する音声生成・知覚を含む音声情報処理に関するワークショップ。

- 期日：1993年11月8日(月)、9日(火)
- 場所：早稲田大学 国際会議場

#### 4.5.4 「不特定話者音声認識」特集号の企画

隠れマルコフモデル (HMM), 神経回路網モデル (neural network model) などを用いた音声認識が活発に研究され不特定話者音声認識に対して高い性能が実現されつつあります。また一方、話者適応により不特定話者音声認識の問題を解決しようとする試みも精力的に行なわれています。誰の声でも認識できる技術の開発は音声認識をさらに普及させるために必須となっています。不特定話者の音声認識に不可欠な音声データベースも充実し、不特定話者音声認識の研究が各研究機関で行なわれております。当分野の研究開発の一層の高揚を図るべく特集の企画。

## 5 むすび

当初筆者が研究室に着任した時にはマイクロ VAX とか言う Workstation (ほとんど X 端末かと思った) と、複数研究員 (複数と言っても多人数の複数) が共同利用する atr-sp (VAX8820) とする超低速のミニコンと、Jstar という訳の分からない高級ワープロと言う、とんでもなく時代遅れで劣悪な環境であった。計算機のレスポンスの悪さや日本語環境の未整備さ、文書処理の未整備さなどに不愉快な思いをしたものであった。国際的な研究所なのでメールはすべて英語でおこなっていたのだそうだが、3年間で計算機環境の整備も進みやっと世間並になり、またある程度研究成果をおさめることができた。今後のさらなる音声情報処理研究・音声言語処理研究の発展を期待したい。

## 謝辞

ATR で3年間を過ごすことという貴重な機会をいただく直接の原因である鹿野清宏博士 (現 NTT HI 研究所) に感謝します。時々激論に対して常に穏やかに対応して頂くとともにのびのびと研究をさせていただいた榊松明社

長に感謝します。研究管理の重要さや忍耐力の大切を教えていただいた嵯峨山茂樹室長に感謝します。研究室運営にあたり多大な助言や援助をいただいた匂坂芳典博士に感謝します。また研究所が異なるにもかかわらず御指導をいただくとともに公私ともにお世話になりました東倉洋一人間情報通信研究所社長に感謝いたします。いろいろ無理難題に対して常に適切な対応をしていただいた研究所企画課の諸氏・諸嬢に感謝いたします。デモ開発、ファイルバックアップ作業、計算機システム管理作業などを滞りなく行ない、研究環境をサポートしてくれた諸氏に心より感謝します。最後に、3年の間、実験や論文執筆において筆者の容赦ない批判と若干の励ましに耐え抜き研究を遂行してくれた部下の諸氏に心よりお礼と感謝をいたします。