

Limited Distribution

TR-I-0350

ATR 自動翻訳電話研究所における 音声情報処理の研究

音声情報処理研究室
(鷹見 淳一・嵯峨山 茂樹 編)

1993 年 3 月

内容梗概

ATR 自動翻訳電話研究所における 7 年間の自動翻訳電話の研究プロジェクトにおいて行なわれた音声情報処理研究の総まとめとして、過去の研究をできるだけ網羅的に記録した。文中には、文献名が略記されており、その略記文献名から、毎年まとめてきた外部発表論文集“Research Activities at Speech Processing Department”中の論文を検索することができる。また、巻末には、現在までに発行された外部発表論文集“Research Activities at Speech Processing Department”の目次を添えてある。

©1993 ATR Interpreting Telephony Research Laboratories

©1993 ATR 自動翻訳電話研究所

目次

1	ATR 自動翻訳電話研究所における音声情報処理の研究の概観	1
1.1	この資料の目的	1
1.2	ATR 自動翻訳電話研究所の音声情報処理研究	2
1.3	音声情報処理研究の概観	2
1.3.1	音声認識の研究	2
1.3.2	音声合成	4
1.3.3	自動翻訳電話実験システム	4
1.3.4	音声データベース	4
2	確率統計モデルによる音素のモデル化の研究	5
2.1	離散分布型 HMM 音素モデル	5
2.1.1	離散分布型 HMM 音素モデルにおける継続時間制御	5
2.1.2	ファジィベクトル符号化離散分布型 HMM 音素モデル	5
2.1.3	複数コードブック離散分布型 HMM 音素モデル	5
2.2	連続分布型 HMM 音素モデル	6
2.2.1	連続分布型と離散分布型の HMM 音素モデル比較	6
2.2.2	混合連続分布型 HMM 音素モデルの改良	6
2.2.3	混合連続分布型 HMM における混合数の自動決定	6
2.2.4	主成分分析を用いた連続出力型 HMM の性能向上	6
2.2.5	識別誤り最少化規準に基づく HMM の学習	6
2.3	音素環境依存 HMM 音素モデル	7
2.3.1	音素環境クラスタリング結果に基づく隠れマルコフモデル	7
2.3.2	隠れマルコフ網	7
2.3.3	時間方向の分割を採り入れた隠れマルコフ網	7
2.3.4	多数話者共通の隠れマルコフ網	8
3	話者適応・不特定話者音声認識の研究	9
3.1	教師あり話者適応を用いた音声認識	9
3.1.1	ベクトル量子化符号帳マッピングによる話者適応	9
3.1.2	ベクトル量子化話者適応の時間遅れ神経回路網による音韻認識への適用	10
3.1.3	話者重畳型 HMM を用いた調音様式の話者適応化	10
3.1.4	移動ベクトル場平滑化を用いた話者適応方式	10
3.1.5	話者混合重みを用いた話者適応方式	12
3.1.6	標準話者の予備選択を併用した話者適応	13
3.1.7	動的な話者適応	14
3.2	教師なし話者適応を用いた音声認識	14
3.2.1	全音素エルゴディック HMM を用いた教師なし話者適応方式	14
3.3	不特定話者音声認識	16
3.3.1	混合連続 HMM による不特定話者音声認識	16

3.3.2	話者混合隠れマルコフ網による不特定話者音声認識	16
4	雑音中の音声認識の研究	18
4.1	雑音環境下における音声認識	18
4.1.1	雑音を除去する方法	18
4.1.2	雑音に対して堅牢な距離尺度を用いる方法	18
4.1.3	雑音環境下音声空間から無雑音音声空間への写像により対処する方法	19
4.1.4	耐雑音性の高いモデルを用いる方法	19
4.1.5	雑音モデルを用いる方法	20
4.1.6	標準音声学習連続分布 HMM の電話音声認識への適用方法	20
4.2	標準音声学習連続分布 HMM の電話音声認識への適用	20
5	LR パーザによる連続音声認識の研究	22
5.1	音素 HMM を用いた連続音声認識	22
5.1.1	HMM-LR 連続音声認識の研究	22
5.1.2	HMM-LR 連続音声認識システムの開発	24
5.1.3	統計的言語モデルに関する研究	24
5.1.4	2段階 LR を用いた文認識の研究	25
5.1.5	到達可能性照合機構を備えた予測 LR パーザの研究	25
5.1.6	音声認識における未知語処理の研究	25
5.1.7	文章発声の連続音声認識に関する研究	26
5.1.8	連続音声認識専用装置を用いた日英自動翻訳システムの構築	26
5.1.9	A* アルゴリズムを用いた探索手法	26
5.1.10	学習可能な探索手法	27
5.2	音素環境依存音素 HMM を用いた連続音声認識	27
5.2.1	音素環境依存 LR パーザの実現アルゴリズムの研究	27
5.2.2	SSS-LR 連続音声認識システムの開発	28
6	言語モデル・韻律利用の研究	29
6.1	言語モデルの研究	29
6.1.1	Bigram 確率統計の利用	29
6.1.2	単語の trigram を用いた連続音声認識	29
6.1.3	単語連鎖可変長統計の自動学習に基づく連続音声認識	30
6.1.4	HMM を用いた文法の自動獲得	30
6.1.5	付属語連鎖、自立語連鎖に基づく言語モデル	30
6.1.6	文脈自由文法の自動獲得	31
6.1.7	行列パーザによる連続音声認識	31
6.1.8	自由発話音声認識	31
6.2	韻律情報の利用の研究	31
6.2.1	韻律の持つ情報量の推定	32
6.2.2	ピッチ周波数パターンを用いた文節区分化	32
6.2.3	超分節的継続時間長制約を用いた音声認識	32
6.2.4	ピッチ周波数パターンを用いた平叙文と疑問文の区別	32
7	ニューラルネットワークによる音声認識の研究	33
7.1	TDNN を用いた音声認識	33
7.1.1	TDNN による音素認識	33
7.1.2	TDNN-LR による連続音声認識	33
7.1.3	TDNN-LR による大語彙単語音声認識	35
7.2	TDNN の改良 (問題点克服の試み)	35

7.2.1	平滑化による頑健性の向上	35
7.2.2	Network 構造の検討	36
7.2.3	対判定型 TDNN(PD-TDNN)	37
7.2.4	音声パターンの時間構造を考慮した TDNN	38
7.2.5	時空間 Block を統合したニューラルネットワーク	38
7.3	確率的出力をもつニューラルネットワーク (Fuzzy Partition Model)	39
7.4	不特定話者音声認識	40
7.5	識別誤り最小規準による学習	41
7.6	堅牢なシステムの構築に向けて	41
7.6.1	FPM-LR による不特定話者連続音声認識	41
7.6.2	ニューラルネットワークを用いた話者独立の特徴抽出	41
7.7	離散値写像ニューラルネットワークと連続値写像ニューラルネットワーク	42
7.7.1	波形上での雑音抑圧	42
7.7.2	音素ニューラルフィルター (PFN)	42
7.7.3	話者適応	43
7.7.4	教師なし構造変換ニューラルネットワーク	43
8	知識処理による音声認識の研究	44
8.1	知識処理に基づく音声認識	44
8.1.1	スペクトログラムリーディング知識に基づく音声認識	44
8.1.2	知識とニューラルネットに基づく音声認識	44
8.1.3	スペクトログラムリーディング知識に基づく自動ラベリング	45
9	諸方式の音声認識性能比較	46
9.1	連続音声認識システム群 ATREUS	46
9.1.1	特定話者文節音声認識の性能比較	47
9.1.2	話者適応文節音声認識の性能比較	48
9.1.3	不特定話者文節音声認識の性能比較	48
9.1.4	特定話者文音声認識の性能比較	49
9.1.5	1500 語、3000 語の語彙の文節音声認識の性能比較	50
10	音声分析、言語識別、音声区分化、その他の研究	52
10.1	音声認識のための音声分析の研究	52
10.1.1	ベクトル時系列のフィルタによる特徴量抽出	52
10.1.2	Dynamic cepstrum の検討	52
10.2	言語識別の研究	52
10.2.1	Bilingual 音声を対象とした日英言語識別	52
10.2.2	スペクトル遷移の確率モデルによる言語識別	53
10.3	音声の区分化・識別問題の研究	53
10.3.1	話者特徴に基づく音声区分化と識別	53
10.3.2	男女性差に基づく音声区分化と識別	53
10.4	その他の研究	53
10.4.1	歪み尺度測地線を用いた音声スペクトルの補間	53
10.4.2	伝送誤りに強い VQ 符号帳構成法に関する検討	53
10.4.3	LPC ケプストラム係数の存在領域について	54
11	音声の規則による合成の研究	55
11.1	非均一合成単位を用いた音声合成方式	55
11.2	統計的手法を用いた韻律制御の最適化	55
11.3	ATR ν Talk 合成システム	56

12 音声変換 (話者変換) の研究	57
12.1 声質変換	57
12.1.1 コードブックマッピングによる声質変換	57
12.1.2 異言語間の声質変換	57
12.1.3 セグメントを単位とする声質変換	58
13 自動翻訳電話	59
13.1 自動翻訳電話 (音声翻訳) の研究	59
13.1.1 自動翻訳電話実験システム SLTRANS	59
13.1.2 自動翻訳電話実験システム SLTRANS2	59
13.1.3 自動翻訳電話実験システム ASURA	61
13.1.4 自動翻訳電話国際実験 (CSTAR 実験)	62
14 音声データベース	63
14.1 音声データベース	63
14.1.1 大規模日本語音声データベース	63
14.1.2 英語音声データベース	63
14.1.3 データベース管理システム	63
14.1.4 音声ワークベンチ	64
14.1.5 各音声データのあらまし	64
14.1.6 データベース販売のご案内	68
15 音声情報処理研究室に在籍した研究員のリスト	72
15.1 研究員リスト	72
16 文中で示した文献のリスト	75
16.1 文献リスト	75

第 1 章

ATR 自動翻訳電話研究所における音声情報処理の研究の概観

1.1 この資料の目的

この資料は、ATR 自動翻訳電話研究所で、1986 年 4 月から 1993 年 3 月までの 7 年間に行なわれた音声情報処理の研究の概観を述べるものである。但し、執筆と編集の時間的な制約から、過去の研究をすべて網羅することはできなかった。また、最近の研究により多くの記述が当てられているのは、研究員の交替などの経緯からやむを得ないことをご了承頂きたい。各部分の分量・正確さ・記述の統一性に大きな差があるが、これは執筆者や情報提供者の協力姿勢が大きく異なるからで、編集者がそれをカバーすることが十分できなかったのが残念である。

この資料中、参照文献名が略記されているが、ここで用いられる文献略記法あるいは識別記号 (identifier) は、ATR 自動翻訳電話研究所 音声情報処理研究室が、過去に外部発表した技術論文をまとめて、以下のような "Research Activities of Speech Processing Department" という名で ATR Technical Report として発行しているものに対応する。

巻	期間	資料番号	論文数	ページ数
Volume I	from April 1986 through December 1988	TR-I-0010	32	181
Volume II	from November 1987 through December 1988	TR-I-0065	88	449
Volume III	from January through October 1989	TR-I-0115	69	282
Volume IV	from November 1989 through December 1990	TR-I-0230	113	574
Volume V	from January through December 1991	TR-I-0261	122	633
Volume VI	from January 1992 through December 1992	TR-I-0349	143	629
Volume VII	from January 1993 through March 1993		~ 42	

それぞれの巻の巻頭には目次があり、論文には、すべて識別記号 (identifier) が付されている。本資料中の "識別記号" のような表記は、関係の深い論文を、上記の論文集の論文識別記号で記したものである。このような意味で、本資料は、上記の資料集へのインデクスとなるものである。

なお、本資料は、以下の人々の調査執筆分担によるものである。また、過去の他の目的の文や資料は最大限活用した。編集は、鷹見 淳一 と 嵯峨山 茂樹が当たった。内容ごとの執筆分担は以下のようである。

- 表紙・前書き：嵯峨山
- 確率統計モデルによる音素のモデル化の研究：鷹見、山口、嵯峨山
- 話者適応・不特定話者音声認識の研究：宮沢、鷹見
- 雑音中の音声認識の研究：大倉、片岸
- LR パーザによる連続音声認識の研究：永井、山口
- 言語モデル・韻律利用の研究：村上、磯谷、Lucke、嵯峨山
- ニューラルネットワークによる音声認識の研究：杉山、福沢、加藤

- 知識処理による音声認識の研究：藤原
- 諸方式の音声認識性能比較：嵯峨山
- 音声分析、言語識別、音声区分化、その他の研究：杉山、嵯峨山
- 音声の規則による合成の研究：三村、岩橋
- 声質変換（話者変換）の研究：岩橋
- 自動翻訳電話：嵯峨山
- 音声データベース：磯谷、嵯峨山
- 研究員リスト：嵯峨山
- 1993年論文リスト：嵯峨山
- 全体・体裁・編集・調整・修正：嵯峨山、鷹見

1.2 ATR 自動翻訳電話研究所の音声情報処理研究

ATR 自動翻訳電話研究所は、1986年4月に発足し、7年間の自動翻訳電話の基礎研究を行なった。本研究は、音声情報処理研究室、データ処理研究室、言語処理研究室の3研究室を擁した。音声情報処理に関する研究は、おもに音声情報処理研究室で、またデータ処理研究室でも行なわれた。本資料は、全体の音声情報処理に関する研究を述べている。

音声情報処理研究室は、以下のような研究項目を推進した。これらの項目は、プロジェクト当初の計画段階に設定されたものであり、プロジェクトを通じて公式の場ではこれらの項目分類が用いられた。

1. 大語彙連続音声認識技術の研究
 - (a) 確率統計的モデルによる音声認識
 - (b) ニューラルネットワークによる音声認識
 - (c) スペクトログラムリーディング知識に基づく音声認識
2. 不特定話者音声認識の研究
3. 会話音声認識の研究
4. 雑音中の音声認識の研究
5. 規則による音声合成の研究
6. 声質変換の研究
7. 音声データベース

1.3 音声情報処理研究の概観

1.3.1 音声認識の研究

音声認識の研究の目的は、自動翻訳電話のための連続音声認識である。その実現のために大きな力が注がれた。

一般に、連続音声認識は、

1. 音声分析（入力音声から音声認識に必要な特徴量を取り出す）
2. 音素モデル（言語を構成する単位である音素の音響的なモデルで、仮定された音素に対し入力音声の確率的な尤もらしさを与える）
3. 言語モデル（事前に仮定される発声内容についての言語的拘束条件で、文法や統計的な性質などを与える）

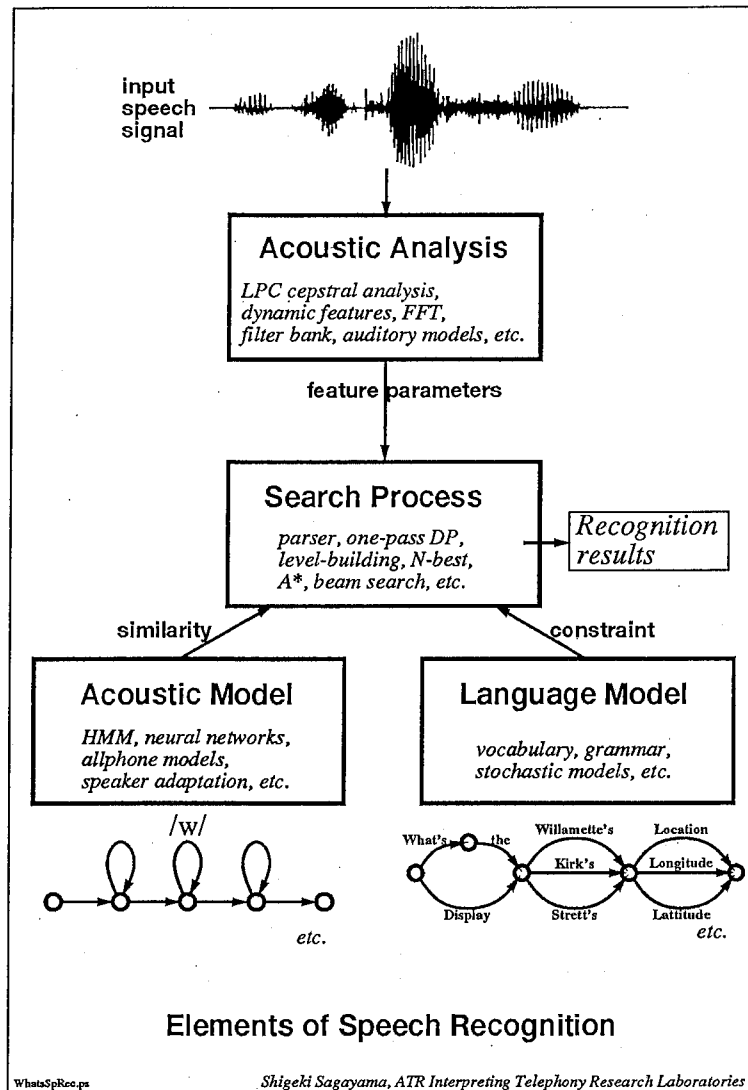


図 1.1: 連続音声認識の概念

4. 探索過程 (入力音声が入素モデル列によりもっともよく説明できるような音素列を言語モデルの制約の中で効率良く探す)

の4つの要素からなると考えて良い。1.1に連続音声認識の概念を図示する。これらの構成要素のそれぞれに重点を置いて各種の研究や検討が行なわれた。

このうち、音素モデルについては、我々は以下の3種類のアプローチを取った。

1. 隠れマルコフモデル (Hidden Markov Model; HMM) などの確率統計的手法
2. ニューラルネットワークによる手法
3. スペクトラムリーディング知識に基づく手法

また、音素モデルについて派生する話者適応・発話様式適応・環境適応・雑音耐性・雑音対策などの手法についても、かなりの検討を行なった。

言語モデルについては、一般化 LR 構文解析を中心として研究を行なった。それに関連して、探索の問題も扱った。言語モデルとしては、統計的な言語モデルや文法の自動獲得などの研究も行なった。

1.3.2 音声合成

音声合成は、規則による音声合成の研究において、自然な音声を生成することを目標に研究を進めた。音声合成の主な構成要素は、スペクトル情報と韻律制御の2つと言える。スペクトル情報に関しては音声合成単位選択法の研究、韻律制御に関してはピッチ・継続時間・パワーの制御の研究などを行なった。

声質変換(話者変換)の研究は、かなり自動翻訳電話に特有のテーマである。自動翻訳電話において、もとの言語の話者の音声の特徴を、翻訳された言語の音声出力に持たせるのが目的であった。このテーマは、音声認識における話者適応と関係が深い。

1.3.3 自動翻訳電話実験システム

音声認識と音声合成と、さらに言語翻訳部を組み込んで自動翻訳電話(音声翻訳)実験システムを構築し、日米独の三か国で自動翻訳電話の公開実験を成功裏に終了することができた。

1.3.4 音声データベース

音声データベースは、研究上不可欠のものである。精力的に開発を続け、現在、好評のうちに外部に販売されている。

第 2 章

確率統計モデルによる音素のモデル化の研究

2.1 離散分布型 HMM 音素モデル

音素モデルの研究において、中心的な検討は、隠れマルコフモデル (Hidden Markov Model; HMM) であった。HMM の型としては、出力分布の型により離散分布型と連続分布型に大別できるが、1986 年よりまず離散分布型とその変形が、そして 1989 年より連続分布型の研究が行なわれた。さらに、1991 年から、HMM の概念の発展である隠れマルコフネットワークの研究を行ない、高い性能を得るに至った。

2.1.1 離散分布型 HMM 音素モデルにおける継続時間制御

ATR における離散分布型 HMM の研究は、1987 年、花沢らによって行なわれた日本語有性破裂音 /b/, /d/, /g/ の認識から始まった。この時点で、語頭と語中のモデルを別に作成したり、学習用音素サンプルに後続母音を 15ms 程度含めて学習するといったノウハウが導入されており、単語音声の 3 音素に対して 92% 前後の認識率が達成されている。[Hanazawa87-10][Hanazawa-87-12][Hanazawa-88-01]

音素または状態の継続時間に制約を与える手法についても検討された。[Hanazawa-88-10a]

2.1.2 ファジィベクトル符号化離散分布型 HMM 音素モデル

離散分布型 HMM では、入力音声はまずベクトル量子化され、離散符号列に変換される。このために、コードブックサイズが小さいと量子化誤差が避けられず、大きいと学習データ量が相対的に不足して統計的に頑健な音素モデルが作れない。このような問題を解決するため、Tseng ら¹のファジィベクトル符号化 HMM の考え方を採用し、量子化を一意決定するのではなく、 k 近傍についてのファジィメンバーシップ関数値の組を用いた言わば「柔らかい」量子化を行なった。これを複数コードブックと組み合わせて用いることにより、離散 HMM 音素モデルの認識性能は大きく向上した。[Hanazawa-88-03] [Hanazawa-88-06]

2.1.3 複数コードブック離散分布型 HMM 音素モデル

音声の特徴量としては、LPC ケプストラムのような音声スペクトルパラメータが良く用いられている。さらに、動的な特徴としていわゆる「デルタケプストラム」を用いると音声認識の性能が向上できることが知られている。これらを統合して一つの特徴ベクトルにまとめると、ベクトルの次元が大きくなり過ぎる。その結果、限られた学習データを用いる場合、コードブックサイズが小さいと量子化誤差が避けられず、大きいと学習データ量が相対的に不足して統計的に頑健な音素モデルが作れない。そこで、これらの特徴量ごとに個別に量子化コードブックを用意することが SPHINX システム²などで行なわれている。この手法を採用することにより、離散 HMM 音素モデルの認識性能は大きく向上した。[Hanazawa-88-10b]

¹ H. P. Tseng, M. J. Sabin, E. A. Lee: "Fuzzy Vector Quantization Applied to Hidden Markov Modeling," Proc. ICASSP87.

² K-F. Lee, H-W Hon, M-Y Hwang, S. Mahajan and R. Reddy: "The SPHINX Speech Recognition System," Proc. ICASSP89, pp. 445-448, 1989.

複数コードブックファジィベクトル符号化離散分布型 HMM 音素モデルは、この後、一般化 LR 構文解析と組み合わせて連続音声認識の手法として ATR において永く用いられ、大語彙連続音声認識の専用ハードウェアの開発、自動翻訳電話実験プロトタイプシステム “SLTRANS” と “SLTRANS2” などに活用されたが、1991 年頃から連続分布型 HMM あるいは隠れマルコフネットワーク (HMnet) によって置き換えられていった。

2.2 連続分布型 HMM 音素モデル

混合連続分布型 HMM は離散分布型 HMM に比べ、より高精度な音素モデルになる可能性があるとともに、話者適応や不特定話者への発展性も期待できることから、1990 年ころから研究を開始し、1991 年から本格的に LR 構文解析と接続して、連続音声認識性能の向上の種々の検討を行なった。現在、標準的な音声認識システム (baseline システム) として、比較の対象や、話者適応のペースにするなど、さまざまに活用されている。

2.2.1 連続分布型と離散分布型の HMM 音素モデル比較

1990 年、花沢 [Hanazawa90-2a] らは日本語子音認識において連続分布型と離散分布型の HMM の比較実験を行なった。連続分布型は高い精度を得ることができるが、発話様式の違いに対するロバスト性では離散分布型の方が有利としている。特定話者の文節発声 18 子音認識の場合、認識率 81.4% を得ている。

2.2.2 混合連続分布型 HMM 音素モデルの改良

その後、山口ら [Yamaguchi92ASJ03a][Yamaguchi92ICSLP10] は混合数を学習サンプル数に応じて決定し、HMM の確率密度関数の分散に対して定数倍操作を施すことにより、発話様式の違いに対するロバスト性を向上させた。特定話者の文節発声 18 子音認識では、92.3% という高い認識率を得た。ここでも離散型と同様、HMM は語頭と語中のモデルを別に作成したり、学習用音素サンプルに後続母音を 15ms 程度含めて学習するといったノウハウが導入されている。しかし、基本的には音素環境独立型のモデルとなっている。

2.2.3 混合連続分布型 HMM における混合数の自動決定

小坂ら [Kosaka92ASJ03][Kosaka92SST12] は学習データの分散の大きさに応じて混合数を自動的に決定する方式を提案し、不特定話者の音素認識実験でその有効性を確認した。ここでの原理は、通常は状態間で「混合数を同数にする」が、「分散を同程度にする」ように混合数を配分する方が性能が上がるというものである。

2.2.4 主成分分析を用いた連続出力型 HMM の性能向上

混合連続型 HMM では、出力確率分布に多次元ガウス分布を用いる場合は、本来は共分散行列は全成分がなければならぬ (いわゆる full covariance) が、実際には学習データ量に限りがある場合は、対角成分のみの多次元ガウス分布とすることが多い。このために、次元間で統計的相関がある場合には、それを無視してしまっている。

このため、学習データを KL 変換によりあらかじめ直交化することが考えられる。また、主成分分析により次元圧縮も考えられる。通常は、直交化はデータ全体に対して行なわれ、一種のパラメータ変換となる。一方、状態ごとに行なうと、full covariance と同等である。そこで、音素カテゴリごとに直交化すると、同一音素内の全 HMM 状態間で tied rotation を行なうことになる。こうして、full covariance とパラメータ直交変換の中間的な統計的 robustness の手法が得られた。[Rainton91SP09][Rainton91ASJ10]

2.2.5 識別誤り最少化規準に基づく HMM の学習

HMM の学習は、多くの場合、尤度最大化の規準に基づいて行なわれる。この他に、さまざまな学習規準が提案されてきた。その中に、認識誤りを最少にする規準が片桐 (ATR 視聴覚機構研究所) らにより提案されている。

連続音声認識の性能を向上させるためにこの原理を用いるには、音素カテゴリ間での識別誤り最少化の規準は不適切であり、単語、文節あるいは文全体の認識率最大でなければならない。このような観点で、適切な規準を用いた学習を提案し、大きな効果があった。[Rainton91ASJ03] [Rainton92ICSLP10] [Rainton92SP01] [Rainton92JASJ06] [Rainton92ASJ10]

2.3 音素環境依存 HMM 音素モデル

高精度な音声認識の実現を目指して、音素音響パターンの変動要因となる音素コンテキストの影響を考慮した、異音 (allophone) を認識の単位とする HMM に関する研究が 1991 年前半から始められた。

2.3.1 音素環境クラスタリング結果に基づく隠れマルコフモデル

音素環境依存 HMM を実現するための初期の手法としては、音素環境クラスタリング (PEC) 手法によって得られる音素環境木の構造を使用し、各音素環境ノード毎に単一ガウス分布 HMM を学習するといった方法により実現された。一般に、音素環境依存 HMM では通常の音素 HMM に比べてモデル数が増加し、有限の学習サンプルによるモデルの学習が大きな問題となる。この手法では、学習サンプルの不足によるモデルの汎化能力の低下を防ぐために、音素環境木の上位クラスタの環境下で学習された HMM を用いてリーフモデルのパラメータを平滑化するという手法が導入されている。この時点で、混合数 15 の混合連続音素 HMM を上回る音素認識率が達成されている。[Takami91ASJ3][Takami91SP6]

2.3.2 隠れマルコフ網

1991 年の後半には、有限の学習サンプルの持つ情報を最大限に活用し、高精度でしかも信頼性の高い音素環境依存 HMM の形成を目指して、状態共有の考え方を取り入れた効率的な音素環境依存 HMM の表現法である隠れマルコフ網 (HMnet) と、その自動生成手法である逐次状態分割法 (SSS) が開発された。[Takami91ASJ10b]

HMnet は、それぞれ固有の異音カテゴリを割り当てることによって独立性を高めた隠れマルコフ状態をネットワーク状に連結したモデルであり、状態共有のない通常の HMM を完全に包含した、より柔軟性の高い HMM の表現方法となっている。

SSS は、出力尤度最大化の基準に基づいて隠れマルコフ状態の分割処理を繰り返すことによって HMnet の精密化を逐次的に行なうアルゴリズムであり、これによって音素環境のカテゴリや状態共有構造、モデルパラメータなどを全て学習処理によって決定する可能となった。

但し、[Takami91ASJ10b] で提案された最初の SSS は、状態の分割が音素環境方向のみに限定されたものであり、一モデル当たりの状態数は事前に与えられていた。

2.3.3 時間方向の分割を採り入れた隠れマルコフ網

SSS はその後も改良が加えられ、[Takami91SP12][Takami92ICASSP3] では、音素環境方向を時間方向に関する逐次状態分割を可能とする SSS が提案された。これによって、一モデル当たりの状態数に関しても自動的に決定することが可能となり、通常の混合連続音素 HMM と比較して少ないモデルパラメータで高い音素認識率が得られる HMnet を生成することが可能となった。[Takami92ASJ3a][Takami92SP9][Takami92JATR]

男性話者一名 (MHT) の音声を用いて SSS により生成した HMnet の 26 音素認識性能を表 2.1 に示す。

表 2.1: HMnet による 26 音素認識実験結果

状態数	総分布数	第1位(第3位まで)の音素誤認識率(%)	
		単語発声データ中の音素	文節発声データ中の音素
200	200	8.98 (0.73)	18.81 (5.12)
400	400	4.70 (0.35)	14.90 (3.62)
600	600	3.15 (0.26)	12.75 (3.86)
800	800	2.42 (0.17)	12.85 (3.81)
1000	1000	2.29 (0.13)	13.04 (4.01)

この HMnet は、その高い認識性能や出力分布として単一ガウス分布を用いた単純なモデルとしての利点を生かして、話者適応手法との組合せによる多数話者音素認識 [Takami92ASJ3b] や音素環境依存型連続音声認識 [Nagai92SP6][Takami92ASJ10a][Takami92SST12][Nagai92ASJ10]、不特定話者音声認識 [Kosaka92SP9] などへと応用され、いずれの場合においても高い認識性能が得られている。

2.3.4 多数話者共通の隠れマルコフ網

SSS はその後もさらに改良が加えられ、1992 年の後半には、HMnet をベースとした話者適応、あるいは不特定話者 HMnet の精度をさらに向上させることをめざして、複数の話者に共通する HMnet の構造を獲得するための手法である 3 領域逐次状態分割法 (3D-SSS) と話者並列逐次状態分割法 (SP-SSS) が開発された。ただし、ここでの実験では、HMnet 生成に必要な計算時間やメモリの都合上、/b/, /d/, /g/, /m/, /n/, /N/ の 6 音素についての HMnet 生成実験しか行なわれておらず、話者共通 HMnet の優位性を確認するまでには至っていない。[Takami92ASJ10b]

第 3 章

話者適応・不特定話者音声認識の研究

自動翻訳電話への入力技術として、不特定の話者が発声した連続音声認識する技術が必要となる。ATRにおいても多数話者の音声データから学習した不特定話者 HMM を用いた研究が行なわれ、ある程度まで不特定話者の音声認識できることが明らかになってきている。しかしながら、音声には、多様な変化が含まれている。よって、単に音声データベースを増やして HMM 音素モデルを学習しても、不特定話者音声認識の問題が解決するとは考えられない。そこで、少量学習資料を用いて話者適応を行なう方法が、不特定の入力話者の音声認識する上で効果的であると考えられ、種々の話者適応方法が ATR において研究されてきた。

話者適応方法は、教師あり話者適応と教師なし話者適応とに大別される。教師あり話者適応は、あらかじめ決められたテキストを発話する必要があるが、少量データで高性能が得られる。一方、教師なし話者適応は、適応化に多くの音声量を必要とする傾向があるものの、発声内容が未知でも学習できるという利点がある。たとえば、認識された音声を認識する場合などのように、認識対象の音声を認識以前に教師なし話者適応に用いることもできる。

3.1 教師あり話者適応を用いた音声認識

ATR における教師あり話者適応方法として、話者間の静的特徴パターンの違いを適応するベクトル量子化コードブックの写像を基本とした話者適応方法、話者間の時系列パターンの違いを適応する話者重畳型 HMM を用いた調音様式の話者適応方法、話者適応を標準話者の音声の特徴空間から入力話者の音声特徴空間への移動として捉えた移動ベクトル場平滑化方式、が研究されその成果が報告されている。また、教師なし話者適応方式では、統計的な言語情報を用いる方法として、全音素エルゴディック HMM を用いた教師なし話者適応方式が提案されその成果が報告されている。以下に各話者適応方式の概要を説明する。

3.1.1 ベクトル量子化符号帳マッピングによる話者適応

ベクトル量子化を用いた符号帳マッピングによる話者適応の手法 [Shikano86ICASSP04] について説明する。ベクトル量子化で、発声者の音声スペクトルを 256 程度のスペクトルからなる符号帳 (codebook) のスペクトル番号で表す。そして、話者適応の問題を、標準話者の符号帳と発声者の符号帳との間の対応関係を求める問題に帰着させることを考える。

LBG アルゴリズム等の符号帳の作成方法によって、標準話者の符号帳を予め用意しておく。新しい発声者 (入力話者) に 100 単語程度発声させ、同サイズの符号帳を作成する。次に、標準話者と発声者が発声した同じ単語音声のスペクトル列を、DP マッチング (動的計画法による時間軸非線形伸縮マッチング) を用いて、両者のスペクトル間の対応付けを行なう。すべての単語について対応関係を求め、この対応付けの分布に基づく平均化操作により、対応する標準話者のスペクトルを決定する。このスペクトルで発声者の符号帳内のスペクトルを置き換える。この手順を数回繰り返す。これにより、発声者と標準話者それぞれの符号帳の中のスペクトル間の対応付けが行なわれる。

さらに、スペクトル、パワー、スペクトル変化の 3 種の符号帳を別々に用意する複数符号帳 [Nakamura88ASE08]、ベクトル量子化歪みを減少させるためにファジィベクトル量子化 [Nakamura89ICASSP05] を用いた。ファジィ

ベクトル量子化は、音声スペクトルに近い数個のスペクトルを符号帳の中から見つけだし、近さの度合い(ファジィベクトル量子化におけるメンバーシップ関数)で重み付けた線形結合により音声スペクトルを近似するものである。この結果、学習に要する単語数も25単語程度に減らすこともできた。

実験結果の一例を示す。ATR音声データベースの4名の発声者(男性3名、女性1名)の文節発声と単語発声を認識実験に用いた。男性1名を標準話者として、HMM音素モデルを5,240単語の音声からForward-Backwardアルゴリズムを用いて作成した。ファジィベクトル量子化による適応化アルゴリズムのための学習用音声として、25単語と100単語を用いた。特定話者文節認識率は、3名(男性2名、女性1名)の平均で88.4%であった。ファジィベクトル量子化による適応化アルゴリズムを用いて、同じ話者に対して文節の認識実験を行なった。この結果、100単語での話者適応では、平均の文節認識率は78.5%、25単語での適応では75.1%となった。話者適応なしでは59.6%であった。

このファジィベクトル量子化による適応化アルゴリズムは、声質変換にも適用することができる。

文献:[Hattori91ASJ3][Hattori91ICASSP]

3.1.2 ベクトル量子化話者適応の時間遅れ神経回路網による音韻認識への適用

ベクトル量子化話者適応アルゴリズムを時間遅れ神経回路網(TDNN)を用いた音韻認識に適用する。TDNNへの適用にあたってはTDNNの入力パラメータの比較、ベクトル量子化を用いた場合のTDNNの構成の検討を行なう。入力パラメータとしては、FFT、LPC分析によるスペクトルとケプストラム、自己相関係数の比較を行なう。また、ベクトル量子化を用いた場合のTDNNの構成としてパラメータ入力、ベクトル量子化の符合を入力とするTDNNを検討し話者適応化を適用する。音韻バランス216単語、重要語5240単語、国際会議申し込みに関する会話文を用いて音声破裂音/b,d,g/の認識実験を男女計4名の話者について行なった。この結果、(1)TDNNの入力パラメータの形式については、周波数領域のパラメータがすぐれている。周波数領域で表現されていれば、FFTでも、LPC分析方法でも大差ない。(2)ベクトル量子化をTDNNに適用する場合、符合列入力形式のTDNNよりもパラメータを入力とするTDNNが優れている。ベクトル量子化としてはファジィベクトル量子化を用いることでベクトル量子化による認識率の劣化は改善できる。(3)ベクトル量子化話者適応アルゴリズムをTDNNに適用した結果、男性間と女性間の平均で78.7%の認識率が得られ、話者適応化が有効に行なえることが明らかとなった。

文献:[Nakamura89SP6]

3.1.3 話者重畳型HMMを用いた調音様式の話者適応化

HMMにおける話者適応化を改善するために複数話者HMM学習を用いた話者適応化を提案する。話者間における発声の違いとしては、(1)静的なスペクトルの違い、(2)調音様式に起因する時系列パターンの違い、がある。本方法では(2)時系列パターンの違いを適応化するために、複数話者HMM学習を用いた話者適応化法を提案する。この方法は、基準標準話者に加えてさらに複数の話者を用意し、基準話者のスペクトルでありながら複数の話者から得られた種々の時系列パターンを1つのモデルの中に含むHMM(話者重畳型HMMと呼ぶ)を生成する。有声破裂音の認識による評価を行なった結果、話者重畳型HMMにより82.5%の認識率が得られ本方式の有効性が確認された。

文献:[Hattori89SP12][Nakamura90IEICEtrans]

3.1.4 移動ベクトル場平滑化を用いた話者適応方式

離散型HMM 移動ベクトル場平滑化話者適応方式

コードブック間の対応付けヒストグラムに基づく話者適応化においては、未知話者の音声空間を十分表現する適応データが必要であり、未知話者の発声した25単語程度の適応データが必要である。本稿では適応データが少なく不完全な対応関係しか得られない場合に、得られている移動ベクトルに対し、ファジィ級関数に基づく内挿および平滑化を行なう移動ベクトル場平滑化話者適応方式を提案する。

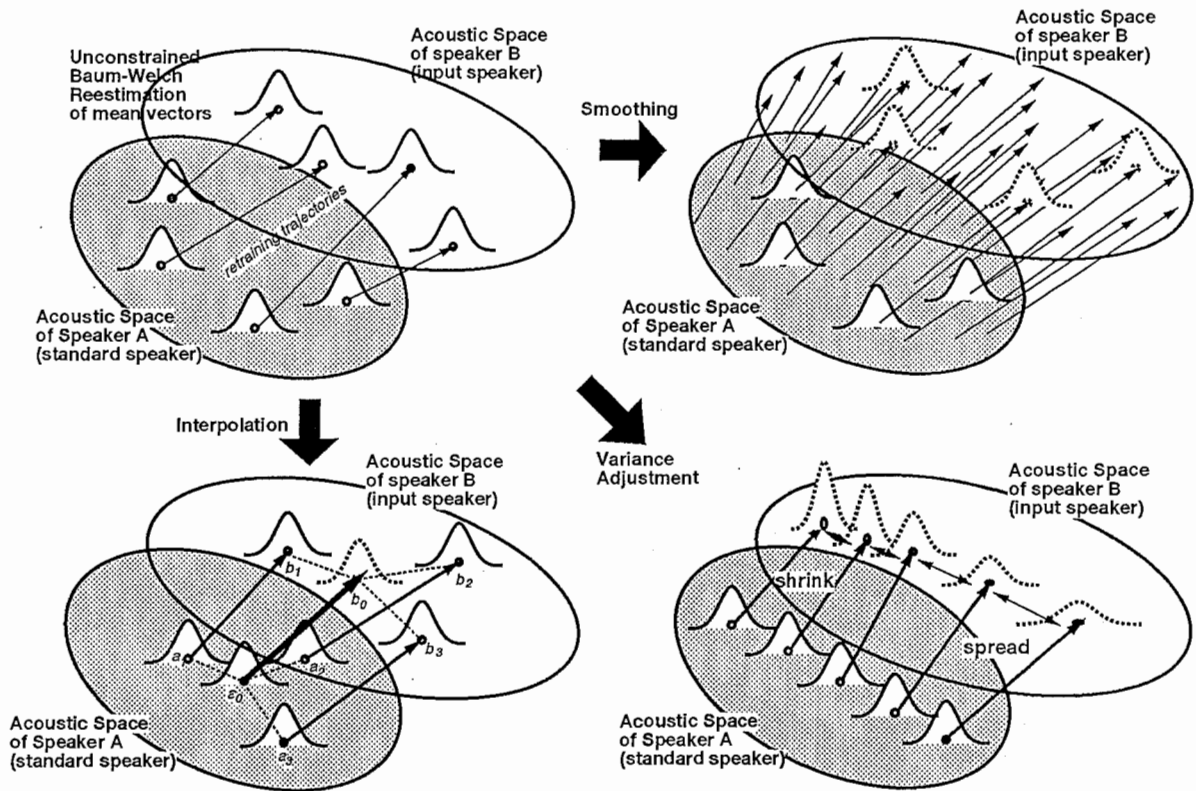


図 3.1: 混合連続分布 HMM 移動ベクトル場平滑化話者適応方式

標準話者として2名の男性話者、入力話者として男女各1名の話者を用いて話者適応を行い、その評価として日本語279文節認識を行なった。平均認識率は、話者適応なしの場合52.3%、従来の対応付けヒストグラムに基づく話者適応化方式で25単語を用いた場合71.0%、提案する方法により5、10、25単語を用いた場合、63.3%、71.0%、73.5%であった。この評価実験結果から、提案する方法により10単語程度の適応データで従来の対応付けヒストグラムに基づく話者適応化方式で25単語を用いた場合とほぼ同等の性能が得られ、本方式の有効性が確認された。

文献:[Hattori92ASJ3][Hattori92ICSLP][Hattori92IEICEtrans]

混合連続分布 HMM 移動ベクトル場平滑化話者適応方式

話者適応の問題を少量学習資料による HMM の再学習の問題として捉えれば、HMM の連結学習を用いて話者適応を行なうことができる。この観点から、移動ベクトル場平滑化を用いた混合連続分布 HMM の話者適応方式である“移動ベクトル場平滑化話者適応方式”を提案する。VFS は、十分に学習された初期モデルの情報を取り込むことにより、不十分な学習資料しか得られない場合の再学習における2つの問題 1) 未学習モデルの存在、2) 学習されたモデルの推定誤差、を解くものである。

本方式は、入力話者の適応音声の発話内容が既知である条件下において、“1. HMM の連結学習”、“2. 移動ベクトルの内挿”と“3. 平滑化”の3ステップより構成され、最終的に平滑化移動ベクトルを用いて話者適応モデルを求めるものである。

まず、標準話者のモデルを入力話者の初期モデルとし、発話内容が既知の適応入力音声を用いて連結学習を行なう。連結学習の結果得られたガウス分布の平均値と初期モデルのガウス分布の平均値との差分を移動ベクトルとする。本方式では、連結学習による平均値の移動を、求めた移動ベクトルに沿った移動と見なす。各平均値に対する移動ベクトルの集合を移動ベクトル場として捉え、移動ベクトルの求まっていない平均値の近傍の移

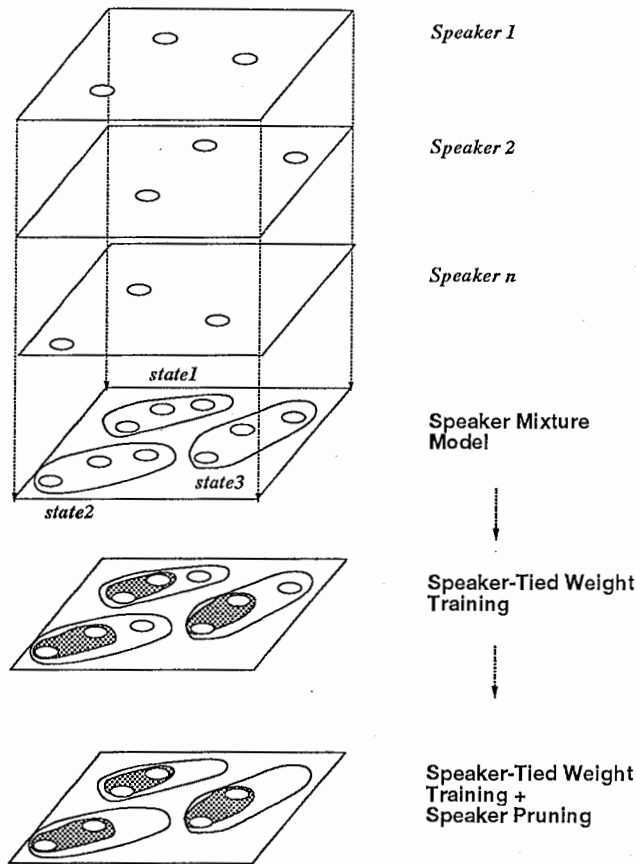


図 3.2: 話者混合 SSS の原理図

動ベクトル場の局所的 / 大局的な方向をもとに、移動ベクトルを内挿する。平滑化も同様に、近傍の移動ベクトル場の方向をもとになされる。[Ohkura92ASJ03b][Ohkura92SP06][Ohkura92ICSLP10][Ohkura92ASJ10b]

VFS は、基本的に少量学習単語時の分布の移動問題を扱うものであり、初期モデルに不特定話者モデルを選ぶこともできる。表 3.1 に、男性 7 名の入力話者が発声した 714 文節の認識実験により、初期モデルに男性話者および不特定話者モデルを初期モデルに用い、適応学習に 50 単語を用いた場合の“VFS” 認識性能を示す。

表 3.1: 入力話者男性 7 名の平均認識率

初期モデル	未適応	平滑化なし	平滑化あり
男性話者	47.7%	74.6%	83.2%
不特定話者	77.1%	76.3%	86.2%

また本手法は、逐次状態分割法 (Successive State Dplitting:SSS) により自動的に生成された音素環境依存モデルである隠れマルコフ網 (HMnet) を用いた話者適応にも応用され、その有効性が確認されている [Takami92ASJ3b]。文献:[Ohkura92ASJ03b][Ohkura92SP06][Ohkura92ICSLP10] [Ohkura92ASJ10b]

3.1.5 話者混合重みを用いた話者適応方式

話者混合逐次状態分割法 (Speaker-Mixture SSS) により作成された不特定話者モデルから、1～5 単語程度の少ないデータで高速に話者適応するためのアルゴリズムとして話者重み学習、さらに話者適応しつつ計算量の減少を行なう「話者プルニング」手法について述べる。

まず、不特定話者の音素環境依存音素モデルである話者混合隠れマルコフ網がすでに作成されているとす

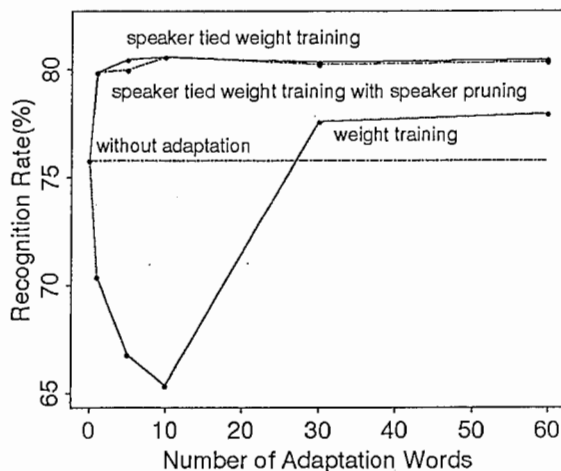


図 3.3: 話者適応後の文節認識実験結果 (200 状態)

る。これについては、不特定話者音声認識を参照のこと。

話者混合 SSS で作成された HMnet の各状態は、各話者に対応する成分を混合して一つの混合分布としたものである。そこで、混合分布の分岐重みを学習することにより話者適応が可能である。学習は通常の Baum-Welch アルゴリズムを用い、出力分布の平均値、分散、遷移確率は更新せず話者重みのみを更新することにより行なう。この学習は連結学習を用いセグメンテーションも同時に行なうため、音素位置のラベルは必要としない。自由パラメータ数は、話者混合の話者数 - 1 であるから、一般に少ない。そのため、わずか 1 ~ 5 単語程度の極めて少ない学習音声により高速に話者適応 (rapid speaker adaptation) ができる。この手法は、話者への分岐確率 (話者重み) を学習するものであるから、いわば尤度最大の原理に基づく話者認識であり、確率的な話者選択でもある。

さらに学習後、話者重みが小さい話者を削除することにより話者適応をしつつ計算量を削減する話者プルーニング (speaker pruning) 手法についても検討を行なった。以上の原理を図 3.2 に示す。また、図 3.3 に話者重み学習による話者適応、話者重み学習 + 話者プルーニング、全重み学習の 3 種類の実験の結果を示す。

文献:[Kosaka92ASJ10b][Kosaka92SP09][Willems92SP12]
[Kosaka93ICASSP][Kosaka93Eurpspeech]

3.1.6 標準話者の予備選択を併用した話者適応

話者適応を行なう際には、入力話者と標準話者との相性の善し悪しが、最終的な話者適応精度に与える影響についても考慮する必要がある。安定して高い話者適応精度を期待するためには、入力話者と相性の良い標準話者を事前を選択する必要がある。

このような背景から、標準話者の予備選択手法についても研究が行なわれた。特に、話者適応後の HMM から得られる出力尤度 (話者正規化 HMM 尤度) を用いる方法は、計算量は大幅に増加するものの、より入力話者に適した標準話者の選択手法として有効であることが音素認識実験 [Miyazawa92ASJ10a]、および文節認識実験 [Takami93ASJ3] によって確認されている。

表 3.2 に、話者適応化 HMnet を用いた連続音声認識における標準話者予備選択の効果を示す (SSS-LR 連続音声認識システムにより評価)。

表 3.2: 話者適応型 SSS-LR 連続音声認識システムにおける標準話者予備選択の効果

標準話者選定条件	第1位(第5位まで)の 平均期待認識率、単位: %
話者選択なし	79.09 (91.96)
性別のみ考慮	86.32 (96.11)
適応前の HMM 尤度 に基づく話者選択	86.38 (96.29)
話者正規化 HMM 尤度 に基づく話者選択	87.38 (96.76)
最適話者の事後選択	88.38 (97.23)

なおこの結果は、標準話者として男性3名および女性3名を使用し、それとは別の男性話者3名および女性話者3名に対してそれぞれの標準話者選定条件に基づいて選択した標準話者を使用した場合に得られた1500語彙の文節認識率の平均を示している。

3.1.7 動的な話者適応

話者適応が音声認識と同時進行し、「話せば話すほど良く認識するようになる」ことは、将来の話者適応の実際的な利用では重要な課題である。このような考え方を「動的話者適応(dynamic speaker adaptation)」と呼ぶことにする。

我々の取ったアプローチは、話者適応速度の異なる複数の話者適応法を並列に使い、これらを切替えて最も高い性能を得ようとするものである。複数の話者適応法としては、たとえば、話者混合重みの学習(全隠れ状態共通の話者重み)、同(全状態独立の話者重み)、話者移動ベクトル場平滑化(VFS)の3種を選ぶ。この順に速度が速いが、学習データ量が多い場合はこの逆順に話者適応性能が高い[Willems92SP12]。

問題は、切替えタイミングが話者によって異なることである。理由の一つは、入力話者に近い話者がたまたま標準話者の中に含まれていれば、話者重み学習の効果が高いため VFS への切替えは遅い方がよい傾向がある、ということである。そこで、入力音声時間で切替えタイミングを切替えるのではなく、3種類の適応を行なって、それらの生み出す尤度により切替えることを行なった[Kosaka93ASJ03]。これは、連続音声認識において、発声内容と話者適応手法の両方を、尤度最大の原理で決定することに相当する。この結果、発声ごとに話者適応法自体が選ばれるために、従来どの単一の手法よりも話者適応性能が高くなった。

3.2 教師なし話者適応を用いた音声認識

3.2.1 全音素エルゴディック HMM を用いた教師なし話者適応方式

教師なし話者適応方式において、統計的な言語情報を用いる方法を提案する。統計的に言語情報を用いるために、言語制約と音響情報とを統合した一つの確率モデルで音声言語を表現することを考える。我々は、そのような確率モデルとして「全音素エルゴディック HMM」を用いる。

全音素エルゴディック HMM とは、図 3.4 に示すように、音素 HMM をある遷移確率によってエルゴディックに連結して構成した、一つの大規模な確率モデルである。確率モデル内の各音素 HMM の各パラメータ(遷移確率、出力確率分布[分岐確率、平均値、共分散行列])の初期値には、標準話者の各音素 HMM のパラメータの値をそれぞれ用いる。確率モデル内の各音素 HMM 間の遷移確率は音素 bigram 確率値に相当するため、なんらかのテキストデータによって求めた音素 bigram 確率値を初期値として用いることができる。このように作成された確率モデルは、言語モデルと音響モデルとの融合したモデルであり、あらゆる言語音声表現するものである。

話者により発声された音声は、音響情報と言語情報とをもつ「言語音声」である。よって、教師なし学習音

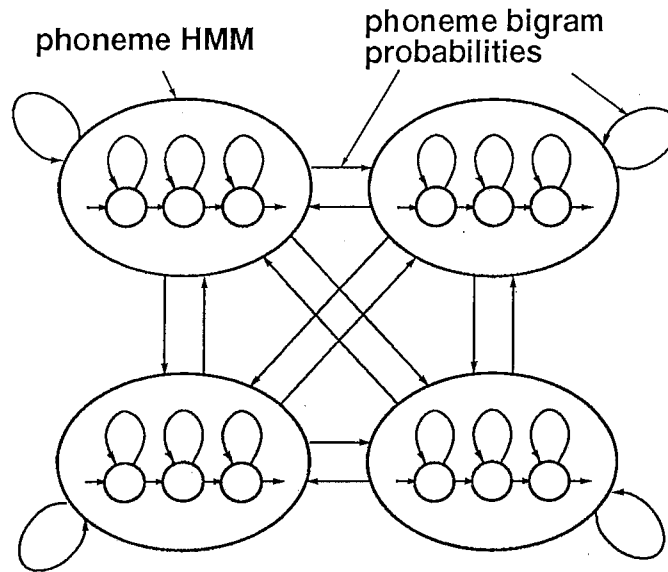


図 3.4: 全音素エルゴディック HMM

声を入力データとして、上記の確率モデルの各パラメータを最尤推定法を用いて学習することが可能となる。但しこの学習では、音素モデルが入力話者に適応されることは理論的には保証できない。たとえば、教師なし学習によって音素モデルがまったく入れ替わることもあり得る。しかし、以下の3つのように学習に制限を設ければある程度この問題に対処した適応学習ができると考えられる。1) 言語モデルの言語的な制約による、音素モデルの連結への制限、2) 入力話者と特性が近い標準話者の音素 HMM による、音素モデルの初期値への制限、3) 移動ベクトル場平滑化による、初期モデルの入力音声空間への移動の制限。これらにより、音素モデルの学習変化は、音素モデルとしての役割の枠の中で学習が進められると考えられる。

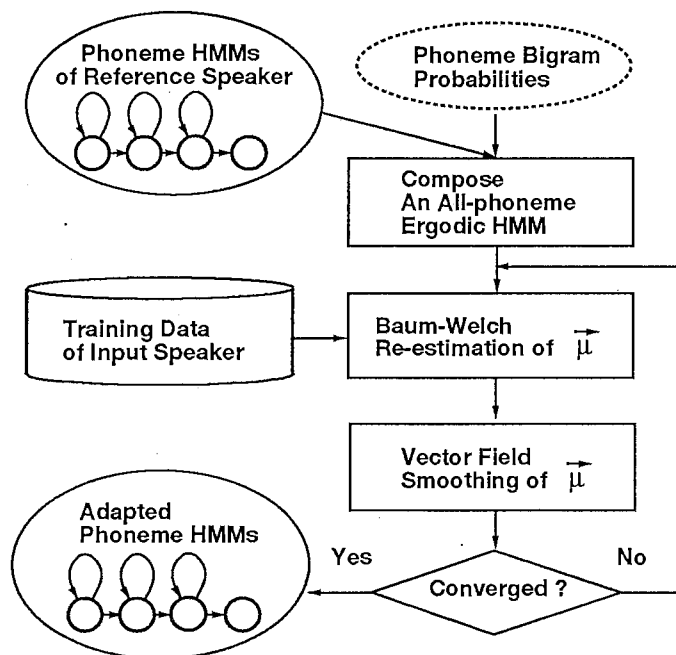


図 3.5: 話者適応学習のブロック図

この学習アルゴリズムのブロック図を図3.5に示す。全音素エルゴディック HMM の各パラメータのうち、適応学習に最も効果のある出力確率分布の平均値 (μ) のみを再学習する。

表3.3に混合数15の不特定話者モデルを初期モデルとして6名の男性入力話者の文節データ (sb3) を認識した結果を示す。比較のために教師あり話者適応[?]の結果と特定話者認識結果も示す。ここで用いた特定話者モデルは音素毎に混合数が2から15まで異なるモデルである。入力話者によっては256文節の適応学習で特定話者モデルの性能を上回った。適応データ量が多い程教師あり話者適応の性能に近づく。

文献:[Miyazawa92ASJ10b][Miyazawa92SP10][Miyazawa93ASJ03]
[Miyazawa93ICASSP04][Miyazawa93FULLPAPER]

表3.3: 初期モデルに不特定話者の混合数15の混合連続型 HMM を用いて話者適応した場合の279文節認識率 (%)

入力話者	話者適応前	教師なし話者適応		教師あり話者適応		特定話者
		102文節	256文節	102文節	256文節	
MAU	75.3	81.0	85.0	87.1	89.6	88.9
MHT	84.8	85.9	89.2	87.4	91.7	87.7
MMS	74.1	80.9	86.3	85.6	86.7	90.7
MNM	82.7	82.4	83.5	89.9	91.0	91.7
MTK	78.4	88.5	89.6	88.5	90.7	91.4
平均	79.1	83.4	86.7	87.7	89.9	90.1

3.3 不特定話者音声認識

3.3.1 混合連続 HMM による不特定話者音声認識

混合連続分布 HMM による不特定話者音声認識については、[Kosaka92ASJ03]において検討が行なわれ、混合数を変えた場合の認識性能の変化などについて報告されている。

3.3.2 話者混合隠れマルコフ網による不特定話者音声認識

話者混合逐次状態分割法 (Speaker-Mixture SSS) により、不特定話者音響モデルを作成する手法を開発した。これは、さらに1~5単語程度の少ないデータで高速に話者適応するためのアルゴリズムとして話者重み学習、さらに話者適応しつつ計算量の減少を行なう「話者プルーニング」につながり、興味深い発展をしている。

従来法の SSS で作られる隠れマルコフ網は、精度の高い音素環境依存のモデルであるが、異話者間の音響的相違が音素環境間の音響的相違より大きいと、異話者間の音響的相違が一種のノイズとなり、不特定話者の音素環境依存のモデルである隠れマルコフ網を作ることは、そのままの SSS アルゴリズムでは難しかった。そこで、単一話者の HMnet を作成し、それを基にして同じ構造 (トポロジー) で別の話者のモデルを多く作成し、最後に状態ごとにそれらを混合することによって混合連続 HMM に似たモデルを作るものである。同じ構造 (トポロジー) の別の話者の HMnet を作成するのは、Baum-Welch アルゴリズムにより再学習すればよいが、そのための学習サンプル量が十分でない場合は、パラメータの学習に話者適応手法 [Takami92ASJ3b] を用い、比較的少ないデータで混合出力分布の学習を可能である。

以下にまとめてアルゴリズムを示す。

初期 HMnet の作成 出力分布として単一ガウス分布を持つ HMnet を、1名の話者データから従来法の SSS により生成する。以下このモデルを初期 HMnet と呼ぶ。

パラメータ学習 次に初期 HMnet に対し話者適応法を利用して複数の話者へそれぞれ適応をおこない、出力分布を変更する。これにより複数話者分の HMnet が作成される。

話者混合化 この複数話者分の HMnet の単一ガウス分布に話者重みをかけ、HMnet の構造中同一位置の状態を1つの混合ガウス分布に表すことにより、複数の HMnet を混合ガウス分布をもつ1つの HMnet に統合する。

話者重みは Baum-Welch アルゴリズムによっても決められるが、予備実験の結果特に効果がなかったため、等確率とした。

文献:[Kosaka92ASJ10b][Kosaka92SP09][Willems92SP12] [Kosaka93ICASSP][Kosaka93Eurpspeech]

第 4 章

雑音中の音声認識の研究

4.1 雑音環境下における音声認識

自動翻訳電話実現のためには、不特定の話者がオフィス等で発声した連続音声を認識する必要がある。話者の問題に対しては、現在、少量学習資料を用いた話者適応により対処している。オフィス等の実環境での問題としては、雑音の問題を解決しなければならない。また、自動翻訳電話実現のためには雑音の問題の他に電話帯域の音声を認識する必要がある。耐雑音性の向上と電話音声認識のために、以下の方法の検討を行なった。

- 雑音を除去する方法。
- 雑音に対して堅牢な距離尺度を用いる方法。
- 雑音環境下音声空間から無雑音音声空間への写像により対処する方法。
- 耐雑音性の高いモデルを用いる方法。
- 雑音モデルを用いる方法。
- 標準音声学習連続分布 HMM の電話音声認識への適用方法。

4.1.1 雑音を除去する方法

文献 [Tamura-88-01], [Tamura88-03], [Tamura88-04], [Tamura88-10], [Tamura-89-01], [Tamura-89-05], [Tamura-89-10], [Tamura90-2], [Tamura90-3], [Tamura90-4] では、波形を入出力とした雑音抑圧ニューラルネットワーク (NRNN) による雑音除去法について述べた。NRNN は、4 層のフィードフォワード型ニューラルネットワークであり、雑音重畳音声を入力し、教師信号として無雑音音声を与えることにより、雑音を除去するためのマッピング関数を実現するものである。本 NRNN の有効性は、スペクトルサブトラクション法とのプレファレンステストによる比較により示した。しかし、NRNN を用いた雑音除去手法では、雑音成分のみでなく音声成分も除去されてしまい、文献 [Ohkura-90-9], [Ohkura91ICASSP5] では、NRNN により起こった音声の歪みをコードブックマッピング手法により補正する手法を提案し、その有効性を示した。SNR が 5dB の環境において NRNN のみを用いた場合の音素 /b,d,g/ 認識率が本手法により 4% 向上した。

4.1.2 雑音に対して堅牢な距離尺度を用いる方法

文献 [Umezaki91IEICEtrans4], [Umezaki91ROC11] では、平滑化群遅延スペクトル距離尺度 (SGDS) の不特定話者音声認識における有効性を示した。

文献 [Ohkura92IEEEWorkshop10] では、雑音に対して堅牢な距離尺度を用いた離散分布型 HMM の耐雑音性向上の検討を行なった。WLR, WGD (Weighted Group Delay), SGDS 距離尺度の比較およびケプストラムのユークリッド距離の比較を行ない、SGDS は WLR とほぼ同等の性能を示すことと、ケプストラムは雑音に対して非常に弱いパラメータであることを示した。また、WGD は距離の正定値性が証明されていない距離尺度であるが、SGD、WLR よりも雑音に対して堅牢である優れた距離尺度であることを確認した。

4.1.3 雑音環境下音声空間から無雑音音声空間への写像により対処する方法

コードブックマッピングによる環境適応

文献 [Ohkura90-6], [Ohkura90-9], [Ohkura91ICASSP5], [Ohkura92IEEEWorkshop10], [Ohkura92ASJ03a] では、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。

本手法は、雑音環境下で作成した未知話者のコードブックを用いてベクトル量子化された未知話者の学習雑音音声と、無雑音環境下で作成した標準話者のコードブックを用いてベクトル量子化された標準話者の学習音声間で非線型時間整合 (DTW) を行うことにより未知話者のコードブック中のコードベクトルと標準話者のコードブック中コードベクトルとの対応付けを行い、対応付けヒストグラムを求めるものである。認識時は、求めたヒストグラムを用いて未知話者の入力コードベクトルを標準話者のコードベクトルに変換することにより、標準話者の HMM を使用して未知話者の音声を認識するものである。また、異空間の写像を可能にするコードブックマッピング手法を用いて雑音の問題のみではなく、話者性の違いを一種の環境の違いと考え、話者環境と雑音環境という 2 つの環境の違いを同時に適応する環境適応の実現について検討し、有効性を示した。

音素 /b, d, g/ を用いた同一話者内における雑音除去実験では、本手法を用いることによりスペクトル歪み、 Δ_{cep} およびパワーの歪みが改善されることが明らかになった。また、文節音声認識実験による異話者間の環境適応実験では、雑音適応および雑音と話者性の違いを同時に適応する環境適応を行うことができることと、適応時と認識時の雑音レベルが異なる場合においても、本手法が耐雑音性をもつことを示した。

移動ベクトル場話者適応手法を用いた雑音適応

文献 [Ohkura92ASJ03a] では、移動ベクトル場話者適応手法 (VFS) を用いた雑音適応方法と上記のコードブックマッピング手法を用いた雑音適応方法との比較を音素 /b, d, g, m, n, N/ 認識実験により評価し、VFS 手法の有効性を示した。

4.1.4 耐雑音性の高いモデルを用いる方法

TDNN と HMM における耐雑音性能の向上

文献 [Ohkura91ASJ3] では、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。

セグメント特徴量を用いた雑音環境下での HMM 音声認識

文献 [Ohkura91SP9], [Ohkura91ASJ10], [Ohkura92IEEEWorkshop10] では、セグメント特徴量を用いた離散分布型 HMM 音声認識 (SQ-HMM) を検討した。

離散分布型 HMM がモデルパラメータ推定用音声資料とコードブック作成用音声資料を分離できることと、コードブック作成用音声資料がモデルパラメータ推定用音声資料よりも少量で済むことを利用し、認識資料から作成したセグメント量子化 (SQ) コードブックを用いて SQ-HMM を再学習し、SQ-HMM を認識資料に適応することにより、セグメントアプローチの発話様式依存性の問題を解決することを試みた。また、発話様式依存 SQ コードブックを用いた SQ-HMM と、雑音に対して堅牢な距離尺度である SGDS (Smoothed Group Delay Spectrum) 距離尺度の組み合わせによる方法の雑音環境下音声認識における有効性を VQ-HMM との比較により検討した。

単語より切り出した18子音認識実験において、SQ-HMMはVQ-HMMよりもSNR=∞、30dB及び20dBの環境において3.9%、8.2%及び9.1%高い認識率を示し、SQ-HMMはSNR=∞、30dB及び20dBの環境において、文節認識実験で88.2%、84.2%、52.7%の認識率を示し、VQ-HMMよりも0.7%、10.0%、11.5%高い認識率を示した。

識別誤り規準を用いた耐雑音 HMM の検討

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。文献 [Ohkura92ASJ10a], [Ohkura93ICASSP04] では、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、MEC 学習を用いた耐雑音の高い HMM の構築方法を述べた。同文献では、学習則 (MEC 学習と最尤推定: Maximum likelihood training (ML)) による HMM の耐雑音性の比較を行ない、複数雑音レベルの音声資料を用いて学習した MEC-HMM は、高い耐雑音性をもつことを示した。

4.1.5 雑音モデルを用いる方法

文献 [Sugiyama92ASJ10] では、LPC ケプストラムベクトルの線形伸縮モデルを一般化し、雑音モデルを用いて雑音レベルを推定することにより雑音抑圧を行なう手法の有効性を示した。

4.2 標準音声学習連続分布 HMM の電話音声認識への適用

文献 [Katagishi93ASJ04] では、マイク入力によって収録された標準音声を用いて学習した混合連続分布 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がマイクのそれとは異なる系からの入力音声の認識手法を提案し、24音素認識実験および文節認識実験により評価した。

ここでの手法の特徴は、HMM のパワー・ケプストラム項の平均値および分散を以下の方法で補正することにある。まず電話音声と標準音声のケプストラムをそれぞれ \hat{c}_n 、 c_n とおくと、電話音声の時不変型線形フィルタ (W) によって作成できれば、 $\hat{c}_n = c_n + h_n$ ($n = 1, 2, \dots, 16$) なる関係が成立し、 $-h_n$ は全極型スペクトル $1/W(\lambda)$ に対する LPC ケプストラム係数として求められる。このことから、ケプストラム項の平均値の補正にはこの h_n を用いることができる。次に、パワー項の平均値とパワー・ケプストラム項の分散の補正に関しては、電話音声と標準音声との差分特徴ベクトルが単一ガウス分布で表現でき、しかも入力音声と無相関であれば、その差分ベクトルの平均値と分散を補正值として用いることができる。表に音素認識性能評価結果を示す。

表 4.1: 音素認識性能評価の比較

方式		認識率 (%)		
		24 音素	子音	母音
標準音声入力		98.0(99.9)	97.5(99.9)	98.4(99.9)
電話 音声 入力	特性補正なし	27.5(80.2)	9.5(59.3)	44.4(99.7)
	ケプストラム項の平均値の補正	62.5(92.3)	46.0(86.6)	78.0(97.5)
	パワー・ケプストラム項の平均値の補正	62.6(92.5)	46.3(87.1)	77.8(97.6)
	ケプストラム項の平均値・分散の補正	74.7(97.2)	61.7(96.5)	86.8(97.8)
	パワー・ケプストラム項の平均値・分散の補正	74.2(96.8)	63.2(96.3)	84.4(97.3)
	$\alpha * \sum_{i,m} (\alpha = 8.5)$	63.0(88.4)	40.1(77.0)	84.3(99.0)
	平均値と $\alpha * \sum_{i,m} (\alpha = 4.5)$	75.0(96.7)	61.7(96.1)	87.4(97.2)
	電話音声で学習した HMM を使用	97.3(99.9)	96.4(99.9)	98.1(99.9)

(): 第5位までの累積認識率

これより、 h_n を用いたケプストラム項の平均値の補正は有効である、 h_n を用いたケプストラム項の平均値の補正と推定値を用いた分散の補正を併用するとさらに有効である、ことが分かる。特に、平均値の補正が有意

であり、分散に関してはその拡大効果から分かるように、一率に定数倍を掛けておけば認識性能はある程度維持できると言える。しかし、標準音声 HMM を補正したものを電話音声認識システムへ適用することによって、音素認識性能 27.5% だったものが 75.0% まで改善されてはいるが、実際電話音声を用いて学習した HMM を用いると、認識性能が 97.3% まで改善されることが表の最下段に示されている。現段階で見るとまだ約 20% の改善効果の差があり、本手法の改良によってさらに改善効果が期待できるものと考えられる。

第5章

LRパーザによる連続音声認識の研究

5.1 音素HMMを用いた連続音声認識

現在、ATRにおけるHMMを用いた連続音声認識システムは、多種多様に非常に多く存在する。それらの大部分は、1988年に誕生したHMM-LR法(北、川端、斉藤)を起源としている。この初期システムは、HMM-LRの原理の提案を主眼として、単一のコードブックの離散型HMMを用いており、音素モデルの高精度化はまだ充分ではなかった(特定話者の1,035語彙タスクに対して、文節認識率が第1位で72.0%、第5位までで95.3%)。

その後、認識性能の向上のために、音響モデル(HMM)とLR構文解析の両面から研究が精力的に進められた。

まず、音響モデルについては、1989年に花沢が、離散型HMMをベースにHMMの高精度化を進め、フアジイ+セバレートベクトル量子化の導入、HMMの状態に対する精密な継続時間長制御などを採用した改良版HMM-LRが構築された。このシステムでは、特定話者で、第1位89.5%、第5位までで99.2%(4名平均)の文節認識率を達成しており、初期システムよりも認識性能を飛躍的に向上させた。この高い認識率は、その後のATR後期における研究の目標となり、1991年の混合連続型HMM-LR(山口)まで、その記録が破られることはなかった。また、この離散VQベースHMM-LRの装置化構想が1989年に始まり、コードブックマッピング方式による話者適応装置(シャープ)と並行して、1989年にHMM-LR連続音声認識装置の設計を三菱電機に発注、1990年に装置の試作を開始、そして、1991年秋に試作装置が完成し、報道発表を行なった。

次に、LR構文解析の研究については、1989年にビームサーチにおける計算量削減法が検討された[Kita89-03]。1990年からは、統計的な言語情報を利用する研究が進められ、音節trigramモデルの導入と確率文法適用の検討がなされた。また、同時期に、文節単位発声の枠組で、構文的制約を予測的に用いて文認識を行なう2段LRパーザが提案された。1991年に入って2段LRパーザはさらに進化し、上位文法カテゴリへの到達可能性照合機構(Category Reachability Test: CRT)が組み込まれ、より効率的な解析が可能になった。一方、1991年には、HMM-LRの大語彙音声認識への適用[Kita91-03-2]や、自動翻訳電話には必須の未知語処理についても研究された。また、LRパーザのSPHINXへの組み込み[Kita91-05-1]も行なわれた。1992年に入ると、文節区切り発声の制約を取り除いた連続発声の文音声認識が研究された。統計的言語モデルとして、確率文脈自由文法と生成規則間のマルコフモデルの二種類のものを用いて、特定話者で文認識率83.9%を達成した。また、探索手法についての研究も進められ、A*サーチアルゴリズムや、ビームサーチの処理量削減のための可変ビームサーチやニューラルネットによる適応型ビームサーチなどが提案された。

以下に、ここで紹介した各研究について概説する。

5.1.1 HMM-LR連続音声認識の研究

音素ベースの隠れマルコフモデル(HMM)とLR構文解析アルゴリズムを統合化したHMM-LRと呼ばれる音声認識方式について研究した。HMM-LR音声認識では、文法から得られる予測的な情報を用いて音声認識の探索空間を縮小するだけでなく、LRパーザの中からHMM音素モデルを直接駆動して認識を行なうために、音声認識と言語処理の間に音素ラティス等の中間的なデータを介する必要がなく、高精度でかつ効率的な大語彙連続音声認識システムを構築することができる。

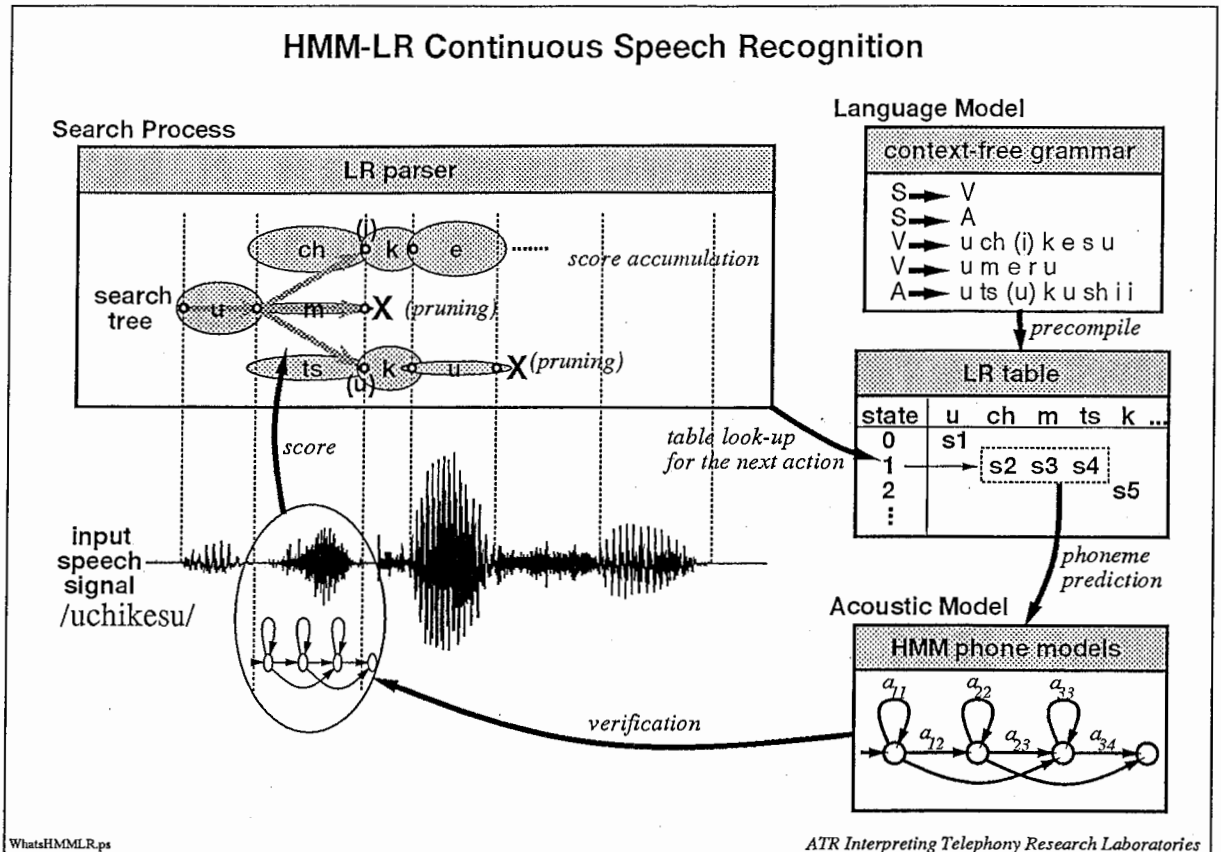


図 5.1: HMM-LR 方式の概念図

HMM-LR 音声認識の動作概略は以下の通りである。

1. 予測 LR パーザにより、その時点までに既に認識された音素系列から文法上で次に接続しうる（文頭であれば文頭にくることが可能な）音素を予測する。音素予測の際に、LR 解析表の現在の状態欄に受理動作が指定されていれば、その音素系列は認識候補として残される。
2. 各音素に対して継続時間長の最小値および最大値等の統計情報があらかじめ求められており、これらの値を使って照合する音声区間を決定する。照合する音声区間は現在までに認識された音素系列と予測された音素の最小継続時間長の和を始端、最大継続時間長の和を終端とする区間である。
3. HMM 音素照合部を駆動し、既に認識された音素系列の音素モデルに予測された音素のモデルを連結して音素照合を行ない、照合スコアを求める。ここで照合スコアとは、照合音声区間が音素モデルから生成される尤度であり、forward アルゴリズムあるいは Viterbi アルゴリズムによって計算される。
4. 照合に成功したすべての音素に対して、並行して音素連鎖の枝を伸ばしていく。実際には音素連鎖の枝を伸ばす過程において解析する候補の数が増加してくるので、照合スコアがある一定値以下の場合は枝刈りするというビームサーチを行ない、解析する候補数を削減する。
5. 再び 1 に戻り、認識を続行する。

最終的に、照合がすべて終わった段階で照合スコアの高い第 n 候補までを認識結果として出力する。

図に、HMM-LR 方式の概念図を示す。

文献: [Kita88-10a] [Kita88-10b]

5.1.2 HMM-LR 連続音声認識システムの開発

HMM-LR 音声認識アルゴリズムに基づいた音声認識システム — HMM-LR 音声認識システム — のインプレメンテーションとシステムの評価を行なった。

最初に、音声の特徴量としてスペクトルだけを用いた単一コードブック型 HMM を使った、日本語の文節認識システムを作成した。このシステムは、特定話者の 1,035 単語を含むタスクに対して、第 1 位で 72.0%、第 5 位までで 95.3% の文節認識率を達成した。

次に、高精度な音韻モデルを用いることにより、システムの改良を行なった。改良システムでは、音声の特徴量として、スペクトル、差分ケプストラム、パワーの 3 種類を用いており、各々別々のコードブックで表現するセパレートベクトル量子化の手法を使っている。また、HMM の各状態の継続時間長の分布を正規分布で表現した継続時間長制御を行っており、これにより高精度の音素認識を実現した。さらに、改良システムにはファジィ VQ マッピングに基づく話者適応化の機能も組み込まれており、新しい話者が数十単語を発声するだけで、システム内部の標準話者の HMM を新しい話者に適応させることができる。改良システムでは、特定話者の場合、文節認識率は第 1 位で 89.5%、第 5 位までの累積認識率は 99.2% (4 人の話者の平均) であり、十分な認識性能を達していると考えられる。話者適応を行なった実験では、第 1 位で 81.6%、第 5 位までで 98.0% の文節認識率を達成した。

混合連続分布型 HMM は離散分布型 HMM に比べ、より高精度な音素モデルになる可能性があるとともに、話者適応や不特定話者への発展性も期待できる。そこで、HMM 音素モデルを従来の離散型から混合連続分布型に置き換えた。混合数を学習サンプル数に応じて決定し、HMM の確率密度関数の分散に対して定数倍操作を施すことなどにより、発話様式の違いに対処した。そして、すべての音素モデルを 4 状態 3 ループにすることで、多大な計算量を要する継続時間長制御を行なわなくても、離散分布型 HMM で継続時間長制御を行なった場合とほぼ同等の認識性能が得られることを示した。なお、HMM の学習に文節発声データを加えると、認識率は飛躍的に向上する (1 名の特定話者認識で第 1 位認識率 88.9% → 93.9%) ことを確認し、同じ発話様式のデータで学習することの重要性も示した。

文献: [Hanazawa89-03] [Hanazawa89-10a] [Hanazawa89-12] [Hanazawa90-04] [Hanazawa90-10] [Kita89-05] [Kita89-06] [Kita89-08] [Kita89-10] [Kita90-03-1] [Kita91-08] [Yamaguchi92ASJ03a] [Yamaguchi92ICSLP10]

5.1.3 統計的言語モデルに関する研究

音声認識では、音響的に似通った数多くの仮説の中から最も確からしい仮説を選択しなければならないが、このためには仮説を定量的に評価する必要がある。通常は、各仮説と入力音声のマッチングのよさを表す音響的な尤度が用いられているが、仮説をより精密に評価するためには、言語的な側面からみた尤度も考慮する必要がある。統計的言語モデルは、仮説の言語的な尤度を計算するのに有効なモデルである。

以下に示す三つの統計的言語モデルについて研究した。

文献: [Kita89-06] [Kita90-03-2] [Kita90-04] [Kita91-05-2]

音節のマルコフモデル

マルコフモデル (bigram, trigram 等) は、非常に多くの音声認識システムで用いられている統計的言語モデルである。これはある単語 (あるいは音素や音節) の後にはどの単語 (音素, 音節) が接続しやすいかという統計情報を用いたモデルである。HMM-LR の拡張である統計的 HMM-LR では、日本語の音節を単位とする 2 重マルコフモデルを採用している。音節の統計情報は、言語データベースの“読み”の部分 (ひらがな文字列) を参照して求めている。

文献: [Hanazawa90-3]

確率文脈自由文法

確率文脈自由文法は文脈自由文法の拡張であり、文法中の各生成規則にその規則の適用される確率が付与されている。確率文脈自由文法では、ある文が導出される確率は、その文の導出に用いられた生成規則に与えられている確率の積として求められる。このモデルは、文法中の各規則がどれだけの頻度で使われるかという統計情報を反映しており、よく使われるような言語表現には高い確率を、まためったに使われないような表現には低い確率を与えるため、音声認識の誤り訂正に有効である。HMM-LR では LR パーザ

を用いて文脈自由文法を制御しているが、文脈自由文法から確率付きの LR 解析表を自動的に構成する方法が知られており、確率付きの LR 解析表を用いることにより効率的に文(あるいは部分文)の生成確率を計算することができる。生成規則の適用確率は、言語データベース中にある文節を実際に解析して、各規則の使われる頻度を調べることにより求めている。

文献：[Kita90-09-2]

生成規則間のマルコフモデル

文脈自由文法を使って、文を解析すると、その文の解析に使われた生成規則の列が得られるが、このモデルはある生成規則の後にはどの生成規則が使われやすいかという統計情報を反映したモデルである。文脈自由文法では、文脈に依存せずに生成規則が適用されるが、このモデルは文脈による生成規則の適用の頻度を考慮したモデルとなっており、文脈依存性を確率の形で持っているといえる。このモデルの統計情報も、確率文脈自由文法と同様に、言語データベース中にある文節を実際に解析してみるにより求めている。

以上述べた3つの統計的言語モデルを HMM-LR 音声認識システムに組み込み、文節認識実験(1名の話者)で評価した。各モデルを用いることにより、第1位での文節認識率は、88.2%からそれぞれ92.5%、92.1%、91.4%に改善された。また、3つのモデルを同時に用いた場合には、93.2%の認識率を達成することができ、モデルの有効性を示すことができた。

5.1.4 2段階 LR を用いた文認識の研究

文節内部の文法的制約と文節間の文法的制約という2つのレベルの文法的制約を段階的に用いる文認識について研究した。またこのために、2段階 LR パーザを開発した。

2段階 LR パーザは、予測 LR パーザの拡張であり、文節間 LR パーザと文節内 LR パーザの2つを用いて、文節カテゴリ(名詞句、動詞句等)の予測と音素の予測を段階的に行ないながら認識処理を進める。この方法では、ある文節カテゴリが予測されたときに、その文節カテゴリに属するような音素系列だけを文節内 LR パーザが生成する必要がある。このために文節内 LR パーザは複数の初期状態を持っており、どの文節カテゴリにはどの初期状態が対応するかという表を LR 解析表に持たせている。2段階 LR アルゴリズムを文節発声による日本語の文の認識に適用した結果、単語認識率95.9%、文認識率84.7%を達成した。

文献：[Kita90-03-2] [Kita90-11] [Kita91-07-1] [Takezawa90-12]

5.1.5 到達可能性照合機構を備えた予測 LR パーザの研究

LR アルゴリズムに上位文法カテゴリへの到達可能性照合機構を組み入れた LR-CRT アルゴリズム (LR Parsing with a Category Reachability Test) について研究した。LR-CRT アルゴリズムを用いることにより、ある特定の文法カテゴリに属する文だけを効率的に求めることができる。このために、LR 解析表を拡張し、各動作項に到達可能な文法カテゴリを持たせた。LR パーザは、移動あるいは還元動作を行なう際に、この拡張された LR 解析表を参照し、これから行なう動作が与えられた文法カテゴリに到達可能であるか否かを調べる。

LR-CRT アルゴリズムを、2段階 LR に基づく文節発声の文認識に適用した。前のセクションで述べた2段階 LR パーザでは、一般に、文節間 LR パーザは複数の文節カテゴリを予測するため、文節内 LR パーザも複数の初期状態を持つことになる。これら複数の状態から同じ音韻が予測された場合、ビームサーチの際に同じ音韻間での競合が起るために、システムの認識性能を落としてしまうという問題がある。LR-CRT アルゴリズムを用いることにより、この問題は解決する。LR-CRT アルゴリズムを用いることにより、単語認識率を97.5%に、また文認識率を91.2%に改善することができた。

文献：[Kita91ASJ10] [Kita92LUC07] [Kita93IEICE01]

5.1.6 音声認識における未知語処理の研究

通常の音声認識システムでは、認識対象となる単語があらかじめ決められていて、これ以外の単語(未知語)は認識することができない。しかし、すべての単語をシステムの語彙中に含めておくのは不可能であり、何らか

の形で未知語処理を行なう必要がある。特に自動翻訳電話では、人名、地名等の固有名詞が頻繁に用いられるので、未知語処理は必須の技術である。

HMM-LR 音声認識システムに未知語処理の機能を組み入れる研究を行なった。基本的な枠組みは HMM-LR と同様であるが、使われる文法が通常のものとは若干異なっている。未知語部分を音素系列として出力するために、タスクを記述する通常文法以外に、日本語の音素間の制約を記述した文法を新たに用いている。音素間の制約を記述した文法では、例えば「子音が連続しない」とか「撥音は語頭にこない」等の日本語の基本的な音素構造が表現されている。この文法の開始記号を、未知語が含まれる可能性のある品詞記号に置き換え、タスクを記述する文法とマージする。このようにして得られた文法を音声認識で使いわけであるが、これだけでは未知語部分の正しい音素系列を得るのは難しいので、音節のマルコフモデルを同時に用いている。また、システムの語彙中の単語が未知語として認識されるのを防ぐために、未知語の含まれている認識候補に対してはペナルティを与えている。

文節認識で使われた文法から固有名詞を取り除いて、固有名詞のみ未知語として認識実験を行なったところ、未知語部分に対して 95.3 % の音素変換率を達成した。

文献：[Kita91-02] [Kita91-03-1] [Kita91-07-2]

5.1.7 文章発声の連続音声認識に関する研究

これまで述べてきた HMM-LR 音声認識システムはすべて文節ごとに発声された音声で認識対象としていた。しかし、文節で区切って発声するというのは、話者に非常に大きな負担を与えることになる。文節発声という制約を取り除き、連続的に発声された文に対する連続音声認識について研究した。

文節発声の場合には単語発声のデータで HMM の音素モデルを学習していたが、連続発声の音声は、文節発声に比べ音素の変形が著しいので、単語発声のデータに加え文節発声および連続発声のデータも HMM 音素モデルの学習時に用いている。音声認識に用いる文法は、文節内文法と文節間文法をマージすることにより文に対する文法を作成し、これを用いている。また、統計的言語モデルとして、確率文脈自由文法と生成規則間のマルコフモデルの二種類のものを用いた。現在、特定話者の場合の文認識率 83.9 % を達成している。

文献：[Kita92ICSLP10] [Kita92SP9] [Kita92AIS04] [Kita92ASJ10]

5.1.8 連続音声認識専用装置を用いた日英自動翻訳システムの構築

1000 語彙の連続音声認識を実時間で処理するために、HMM-LR 連続音声認識装置を設計、開発した。HMM-LR 法による実時間の連続音声認識を達成するために、本装置はパイプライン処理、並列処理及び分散処理等の種々の処理高速化手法を採用した。その結果、連続文節発声の入力音声に対し、入力文の長さによらずに、発声終了後から 2～3 秒で全ての認識処理を終了する処理速度を達成した。

また、HMM-LR 連続音声認識装置、及び話者適応化装置を組み込んだ、高速の日英自動翻訳デモシステムを構築し、音声入力から翻訳結果出力まで 10～20 秒程度の処理時間を達成した。

文献：[Nagai91ASJ10b] [Nakamura91ASJ10] [Nagai92ICSLP10-1]

5.1.9 A* アルゴリズムを用いた探索手法

混合連続分布型 HMM を用いた高精度な音素モデルを HMM-LR 連続音声認識に組み込み、特定話者の文節認識実験を行ない、良好な結果を得ている [Yamaguchi92ASJ03a]。しかし、現在用いているビームサーチは最適性が保証されていない上、高い認識率を得るためにはビーム幅を大きくしなければならず、多くの音素照合回数を必要としていた。

今回新たに A* アルゴリズムを導入した。われわれは計算の効率を考慮し、すべての音素モデル同士が接続可能として、未探索部分の入力音声の対数尤度の推定値: \hat{h}_n とし、探索とは逆方向に、文の終端から始端に向かって Viterbi Algorithm で求めた。ただしこのままでは認容可能性は満たしているものの、ヒューリスティック力が弱く盲目的探索に陥る危険がある。そこで、現時点のノード n から LR テーブルを 1 つ先読みして、そのノードに接続可能な音素候補の集合を求め、音素モデル同士の接続に制限を設けた。さらに、 $\hat{h}_n(t)$ に探索が深くなるにつれて寄与が小さくなるように動的重み付けをして最適条件を緩和した [?]

混合連続分布型 HMM-LR を用いて文節認識実験を行ない、ビームサーチ、best-first 探索と認識性能、計算量について比較した。認識率の観点からは従来からのビームサーチが最も優れていたが、大きな探索空間を必要とする。A* アルゴリズムは音素認識が良好に動作すると探索空間をビームサーチの 1/10 以下に減らすことができ、処理の高速化が期待できる。今後音素モデルの改良が進み精度が向上すれば、A* アルゴリズムは探索方式の有力な候補となりうる。

文献：[Yamaguchi92ASJ03a] [Yamaguchi92ASJ03b] [Yamaguchi92ICSLP10]

A. Barr, E. A. Feigenbaum 著 (田中、淵監訳)：“人工知能ハンドブック”，第 II 章，共立出版 (1983)。

5.1.10 学習可能な探索手法

音声認識システムではこれまで一般に、ビームサーチに代表されるヒューリスティックな探索手法を採用して来た。これらの探索手法は最適な探索となるために、多くの音素照合回数、つまり大きな探索空間を必要とする。そこで、われわれはこの音素照合回数を削減し、認識処理の速度を上げるため、ニューラルネットワークを用いて制御する学習可能な適応的探索手法を開発した。

本方式は各候補の得点などの観測可能な特徴量を入力とするニューラルネットワークからなる制御関数を用いて、探索範囲を適応的に変化させることを特徴とする。この制御関数は学習により得られているので、タスクや音素モデルの性能に応じてビーム幅の制御を最適に行なえる。混合連続分布型 HMM-LR を用いて従来法との文節認識比較実験を行なったところ、提案した適応的ビーム探索は認識率の若干の低下を伴うものの、音素照合回数は平均して 1/3 以下に減少した。

また、この方式は制御関数を学習し直すことにより、異なる問題領域に対処可能である。つまり探索アルゴリズム自体はを交えることなく、同じアルゴリズムで様々な問題に自動的に対応できるという汎用性に富んだ探索方式である。

文献：[Yamaguchi92ASJ10a] [Yamaguchi93IEICE01] [Yamaguchi93ICASSP04]

5.2 音素環境依存音素 HMM を用いた連続音声認識

高精度の連続音声認識を実現するためには、音素環境による音素パターン変動の情報を利用して、認識単位として音素環境依存の異音モデルを用いることが有効である。このような音素環境依存の音素モデルを活用して連続音声認識を行なうには、音素環境に依存した構文解析手法が必要である。このために、音素変形要因の中で特に重要な音素コンテキストを考慮し、LR 構文解析の枠組で、音素コンテキスト依存の LR 構文解析の実現手法が三方式研究された。この音素環境依存 LR 構文解析法は、最終的に逐次状態分割法 (Successive State Splitting: SSS) により生成された隠れマルコフ網 (HMnet) と組み合わせられ、ATR において最高認識性能を有する SSS-LR 連続音声認識システムが開発された。

5.2.1 音素環境依存 LR パーザの実現アルゴリズムの研究

先行、中心、後続音素の 3 要因音素コンテキストに依存した構文解析を行なう LR パーザのアルゴリズムが三方式、研究された [Nagai93IEICE01]。

一般に、音素コンテキスト依存の音素モデルを駆動する場合、単語の両端の音素のコンテキストは不確定であり、音素コンテキストによってどのように単語間の接続を行えばよいかという問題を単語間接続問題 (word juncture problem) と呼ぶ。以下、順に開発された三方式は、いずれもこの単語間接続問題を解決している。

1. パーザレベルでの実現

音素コンテキストに非依存の LR テーブルを用いて、パージング中に予測音素の先行音素、後続音素を動的に参照し、音素コンテキストを予測する方式。まず、予測音素を中心音素とし、先行音素情報は構文解析の履歴を参照し、後続音素情報は次に遷移する予定の LR 状態を参照して、中心音素の音素コンテキストを予測する。次に、予測された音素コンテキストに適合する異音モデルを用いて音素照合計算を行なう。

文献: [Nagai90-9] [Nagai91SP6]

2. テーブルレベルでの実現

LR テーブルを変換して、音素コンテキスト依存の構文解析動作が可能な LR テーブルを生成する方式。まず、文脈自由文法から LR テーブルを構成する段階で、LR テーブル中の各予測音素に対する、出現可能な音素コンテキストを決定し、次に、出現可能な異音について、異音同士が接続し得るかどうかを判定しながら、LR テーブルの状態を分割操作して、正しい異音間の接続を生成していくアルゴリズム。

文献: [Nagai91ASJ3] [Nagai91SP6] [Nagai91Eurospeech9]

3. 文法レベルでの実現

文脈自由文法を変換して、音素コンテキスト依存の文脈自由文法を生成する方式。まず、文法を再帰的に探索して、各非終端記号ごとに出現可能な音素コンテキストを決定し、非終端記号の語頭、語尾で出現可能な異音を決定して、非終端記号にユニークな名前を付与する。次に、非終端記号同士の接続が可能かどうかを判定しながら、新しい規則を生成していくアルゴリズム。LR 構文解析に限らず、文脈自由文法を扱う枠組のパーザに適用可能。

文献: [Nagai92ASJ3] [Nagai93IEICE01]

以上の音素コンテキスト依存 LR パーザを用いて、音素環境クラスタリング (Phoneme Environment Clustering: PEC) により抽出された異音モデルを駆動する、PEC-LR がそれぞれ三方式について開発された。文節認識実験による性能評価の結果、音素環境依存による認識性能の向上が確認された。しかし同時に、量的制限のある音声データを用いてモデルを推定しようとする場合、異音数の増加に伴い学習サンプル数が減少し、モデルの推定精度が低下するという問題点も明らかになった。

文献: [Nagai91ASJ10a]

5.2.2 SSS-LR 連続音声認識システムの開発

前節の問題を解決するために、モデルの中からできる限り無駄な表現を省き、必要最小限のモデルパラメータで学習サンプルの持つ情報を効率良く吸収する学習アルゴリズムの逐次状態分割法 (Successive State Splitting: SSS) が提案された。そして、SSS によって生成される隠れマルコフ網 (Hidden Markov Network: HMnet) の高精度な認識性能が、種々の音素認識実験を通して示されてきた。

そこで、このような高い認識性能を有する HMnet を駆動するべく、SSS と音素コンテキスト依存 LR パーザが組み合わせられ、SSS-LR 連続音声認識システムが開発された。音素継続時間制御にも SSS の原理が生かされ、SSS を用いて HMnet とは独立に生成した音素コンテキスト依存の音素継続時間モデルが採用された。1000 語彙タスク (文法規則数 1407) の特定話者 (MAU) の文節認識実験を行なった結果、第 1 位で 93.2% の高い認識率を得、Fuzzy VQ に基づく離散 HMM や混合連続分布 HMM と比較して認識誤り率が約 4 割減少した。更に、ビーム幅を 32 まで絞っても第 1 位で 89.3% の高い認識結果を得た。これより、高性能で高速の連続音声認識システムの構築が達成された。

その後、ベクトル場平滑化手法 (VFS) による話者適応により、多数話者に対する評価が進められ、1649 語彙タスク (文法規則数 2809) の 4 名の話者 (男 2 名、女 2 名) に対して、第 1 位で平均 90.2% の文節認識率を得た。また、3702 語彙にタスクを拡張した場合の特定話者についての文節認識率は、第 1 位で、89.8% であった。話者適応込みの評価では、4 名の話者 (男 2 名、女 2 名) に対して、第 1 位で平均 87.2% の文節認識率を得た。

さらに、連続発声の文音声に対する認識性能も評価された。文節内文法と文節間文法を一つの文脈自由文法に統合した文認識用文法 (744 語彙) を用い、ファジィ VQ による離散 HMM を用いた結果と比較して、文認識率で第 1 位が約 6.5% (57.7% → 64.2%) の向上が確認された。なお、SSS-LR は、ATR 自動翻訳電話実験システム ASURA の音声入力部として用いられ、日英独三ヶ国間自動翻訳電話の公開実験において重要な役割を果たした [Sagayama93ASJ3]、[Takezawa93IEICE]。

文献: [Nagai92SP6] [Nagai92ASJ10] [Takami92ASJ10] [Nagai92ICSLP10-2]

第 6 章

言語モデル・韻律利用の研究

6.1 言語モデルの研究

この章では、音声認識のための言語モデルの研究について述べる。連続音声認識は、音声分析、音素モデル、言語モデル、探索過程(パーザ)の4要素からなると言って良い。従って、言語モデルと探索過程(パーザ)の研究は、音素モデルの研究同様に重要である。ATR においては、文脈自由文法(CFG)を言語モデルとし、一般化 LR 構文解析法をパーザとする方式を広く用い、確率的音素モデルとも、ニューラルネット音素モデルとも、組み合わせて良好な連続音声認識を達成した。しかし、一方では、他の言語モデルやパーザの研究も行ない、さらには言語モデルの自動獲得の研究も行なった。この研究項目を以下にあげる。

1. Bigram
2. Trigram
3. 単語連鎖可変長統計
4. HMM
5. 付属語連鎖、自立語連鎖に基づく言語モデル
6. 文脈自由文法の自動獲得
7. 行列パーザによる連続音声認識
8. 自由発話音声認識

6.1.1 Bigram 確率統計の利用

[Murakami92ASJ03]では、パーザとしては one-pass DP を用いて、単語 bigram 統計を用いた連続音声認識を行なった。これは、trigram を用いるパーザアルゴリズムとの比較において baseline として用いた。

6.1.2 単語の trigram を用いた連続音声認識

連続音声認識において、認識性能を向上させるため、言語情報が使われている。この中で、bigram や trigram などの Markov モデルが、一般的に利用されている。しかし、単語のみの trigram モデルを用いた連続音声認識アルゴリズムは、認識時に多くのメモリと計算量が必要であるため、品詞と単語を組み合わせた trigram によって、これらの問題を解決している。[Murakami92ASJ10b]では、単語の trigram を用いた連続音声認識アルゴリズムについて報告した。このアルゴリズムは One-pass DP をベースに、一種の 2 次元的な DP 計算を行なうものである。問題となるメモリ空間や計算量は、ビームサーチおよび Viterbi の経路計算方法を改良することにより、大幅に削減できた。そして認識実験の結果、テキスト open では 40%、テキスト closed では 76% の文認識率が得られた。このアルゴリズムでは、計算量は、ビーム幅に大きく依存し、言語モデルには、あまり

依存しない。従って、テキスト closed の認識に限れば、言語モデルの Markov モデルの次数を上げることにより、認識性能は向上できる。従って、今後の研究として、任意のテキストにおいて、高いカバー率と低い perplexity を持った言語モデルの研究が必要であると思われる。

6.1.3 単語連鎖可変長統計の自動学習に基づく連続音声認識

連続音声認識システムにおいて、認識性能を向上させるために、言語情報が利用されている。そして、経験的には、perplexity と文認識率には、負の相関があることが知られている。したがって言語モデルとしては perplexity が低く、かつカバー率が高いものが望ましい。言語モデルとして単語連鎖統計 (n -gram モデル) を用いる場合、低い perplexity を追求すると、bigram よりも trigram さらに 4-gram などの言語モデルを選択できるが、逆にカバー率は低くなる。この問題を解決するために、認識する単語によって bigram, trigram や 4-gram などのマルコフモデルを使い分ける方法 (単語連鎖可変長統計) がある。この選択方法には、様々な基準が考えられるが、言語の perplexity を下げるように、選択する方法がもっとも好ましいと予測される。

[Murakami92ASJ10a] では、単語連鎖可変長統計の選択方法、およびこのモデルを使用した認識実験の結果について報告した。なお、同様な方法が提案されているが、これらは、実際の音声認識実験をおこなっていないため、本稿の方法と、言語モデルの生成方法が異なっている。得られた単語連鎖可変長統計モデルの解析した結果、このモデルは日本語の言語の特性を良く表現していることが確認された。また、認識実験から、bigram モデルとほぼ等しいノード数では、テキスト closed の条件では、bigram モデルとほぼ同程度の性能が得られた。テキスト open の実験では、bigram より、やや低い性能しか得られなかったが、ノード数を増加させることにより、認識性能は向上する傾向が見られた。なお、本稿で述べた学習アルゴリズムは、bigram と比較すると、同等のノード数において、エントロピーは高いため、他の学習アルゴリズムを考慮する必要があると思われる。

6.1.4 HMM を用いた文法の自動獲得

音声認識に用いられる言語モデルとしては、ネットワーク文法 (有限状態オートマトン) や文脈自由文法など構文情報を記述した構文的モデルと、bigram, trigram など統計的な情報を記述した統計的モデルなどがある。また、これらの両方の特徴を持つ言語モデルとして、確率つきネットワーク文法が用いられている。しかし、従来の確率つきネットワーク文法は、人間が言語知識に基づいて記述しているため、大規模な言語現象のモデル化に労力がかかるなどの問題がある。そこで [Yamamoto92ASJ10] では、HMM が、ネットワーク文法と構造が類似しており、かつ Baum-Welch アルゴリズムにより学習できることに着目して、単語を出力シンボルとする ergodic HMM によって文をモデル化することにより、言語データから確率つきネットワーク文法を自動的に獲得することを試みた。HMM による言語のモデル化の研究は、村瀬、田本等によって報告されているが、これらは単語を品詞など複数のカテゴリーに分類したものの連鎖をモデル化するものであった。これに対して、本方式では、単語そのものの連鎖を HMM でモデル化する。これにより、状態遷移出力の偏りとして、単語のカテゴリーも同時に学習されることが期待できる。さらに、得られた ergodic HMM を言語モデルとして連続音声認識に用いて、bigram を用いた場合との認識結果の比較を行なった。実験の結果、出力確率の学習結果の解析から品詞概念の自動獲得がなされていること、遷移確率の学習結果の解析から品詞間の接続関係のような文法が獲得されていることが分かった。このように自動獲得された言語モデルを perplexity、文の出力確率、音声認識性能などにより検討したところ、従来の bigram より若干よい結果が得られた。

[Yamamoto92ASJ10][Yamamoto92IPS10]

ただし、データ量、HMM 状態数ともまだ少ないので、今後も検討を続ける必要がある。

6.1.5 付属語連鎖、自立語連鎖に基づく言語モデル

音声認識に用いるための新しい統計的言語モデルとして、文中の付属語のみに注目した N -gram、自立語のみに注目した N -gram を提案した。これは、単語 N -gram を発展させ、より大域的な関係を表現できるようにしたものである。文中の文節末にあらわれる助詞などの付属語のみに注目してその連鎖を統計的にモデル化することにより、構文的な関係に相当する情報を抽出できると考えられる。同様に、自立語の連鎖により意味的な関

係を表現できると期待される。大量のテキストデータから実際にパラメータを学習することにより、日本語における語と語の間の構文的、意味的な関係をとらえられることがわかった。また、音声認識への応用の簡単な例として、bigramの場合について文節ラティスから文候補を選択する実験を行なった。その結果、文候補の選択にCFGで書かれた文節間文法を用いる場合より高い性能が得られ、本モデルが音声認識の性能向上に有効であることを確認した [Isotani93ASJ03]。

6.1.6 文脈自由文法の自動獲得

Inside-outside アルゴリズムは通常文全体、あるいは文節などの文法を自動獲得することに用いられるが、それでは一般にデータ量が十分は得られない。そこで、単語部分列に適用して、単語部分列の文法記述とその最適分割を行なった [Lucke92ASJ10]。

An algorithm is under development that is capable of inferring stochastic context-free grammar rules from example text. The algorithm is designed to read a continuous source of text, segment it into small units and construct parse trees over these segments. The size of the segments depends on the modelling capacity of the grammar and generally grows during training. The training algorithm is based Bayesian belief networks, and the parameters involved may be optimized using the E-M algorithm. The algorithm can be shown to converge. It is completely unsupervised, requiring as input only an unlabelled source of symbols (words). Moreover, since the model is based on a mechanism dealing with uncertainty, it interfaces in a natural way with a speech recognizer that outputs lists of candidate words together with their likelihood scores.

6.1.7 行列パーザによる連続音声認識

HMMによる確率計算法として、行列計算による手法を考案した。
[Sagayama90SP12][Sagayama91Eurospeech10]。

これは、HMMの状態群について、その中にフレーム時刻 i に遷移して入り、フレーム時刻 j に他に遷移する確率 p_{ij} を (i, j) 成分とするような行列を扱うHMMの計算法である。こうすると、直列の経路は行列の積で、並列の経路は行列の和で簡潔に表せる。理論的に興味深いだけでなく、forward アルゴリズムに等価な計算が可能であり、継続時間制御が容易で、文脈自由文法とHMMを統合した形式で扱えるなどの特徴がある。TDNNとの組合せ [Singer92SST12] や、HMMのモーラ継続時間制約の利用 [Singer93ICASSP04] などが検討されている。

6.1.8 自由発話音声認識

自由発話の音声認識は今後の大きな課題であるが、従来の朗読発声の認識と比較するとさまざまな困難が予想される。自由発話の研究としては自然言語処理の立場から行なわれた研究や、英語における予備的な報告などがあるが、[Murakami92SPREC02]では自由発話の認識にむけて自由発話を音響的な面から融合ラベルの頻度と音素認識率、言語的な面から冗長語および言い誤り・言い直しについて調査した。この結果朗読音声と比較すると融合ラベルの出現頻度は約1.5倍に増加すること、また音素認識率はかなり低下することが示された。そして従来の文法の枠組では冗長語や、言い誤りを考慮していないために自由発話全体の約5割のみをカバーしていることが示された。これらの対策としては、前者においては音素の認識性能を向上させるために、より高いロバスト性をもった認識系を考える必要がある。後者に対しては、冗長語を採り入れた文法をつくることにより9割の文章をカバーできる可能性があるが、言い誤り、言い直しなどの言語現象があるため、基本的に言語処理系を考え直す必要があると思われる。

6.2 韻律情報の利用の研究

韻律情報が音声認識において役に立ちそうであることは何度も指摘されている。しかし、実際の利用は容易ではない。特に、現在までに開発されてきた効率の良い連続音声認識のアルゴリズムに、韻律情報の扱いを含め

るには、まだ多くの問題がある。

ATR においては、韻律情報の利用に関していくつかの検討を行なった。それらは、韻律の持つ情報量の推定、ピッチ周波数依存の音素モデル (HMM の章を参照)、ピッチ周波数パターンを用いた文節区分化、超分節的継続時間長制約を用いた音声認識、ピッチ周波数パターンを用いた平叙文と疑問文の区別、などである。

6.2.1 韻律の持つ情報量の推定

韻律を音声認識に利用する前に、韻律情報がどれほどの情報量をもつか定量的に調べられることが望ましい。これは、音声認識へ韻律を利用する場合の性能向上の上限を規定するものとなる。

これを、近似的ではあるが、定量的に調べるために、かな漢字変換と漢字かな変換を用いて、かな文字列に韻律情報 (ここではアクセント句境界位置とアクセント核位置) を与えた場合に対応する漢字列候補がいくつ生じるか、与えない場合にいくつ生じるか、の数の比により、曖昧さ解消の情報量を求めた [Murakami91SP9]。その結果、文あたりのアクセント句境界位置とアクセント核位置の持つ情報は 5.16 ビットで、かな文字 1 字の情報量に近いことが分かった。なお、この研究はおもに NTT において行なわれ、ATR にて理論化や論文化が進められたものである。

6.2.2 ピッチ周波数パターンを用いた文節区分化

ピッチ周波数パターンを用いて、文音声をアクセント句 (ほぼ文節に対応する) へ分割することを試みた [Shimodaira90-12]。これは、文音声をまず文節に区切って、文節ごとに文節文法のもとで認識をするような場合に利用できる。

文音声は、いくつかのアクセント句の接続であると考えられる。そこで、まず、さまざまなアクセント句のピッチ周波数パターンをクラスタリングして少数のアクセント句ピッチ周波数パターンを作成する。そして、文音声のピッチ周波数パターンを、連続単語認識のアルゴリズムである one-pass DP アルゴリズムにより、どのようなアクセント句ピッチ周波数パターンの接続であるかを決定する。同時に、trace-back (あるいは Viterbi decoding) によりアクセント句境界が決定できる。

この考え方は、ピッチ周波数パターンについての音声学的な知識や理論や学説を極力排して、単純にパターン認識のアプローチを用いたものである。

6.2.3 超分節的継続時間長制約を用いた音声認識

韻律の他の面として、時間特性がある。

HMM では状態継続時間の制約は単純ではない。しかし、種々の方法により、状態継続時間に制約を設け、そのように制御することで音声認識性能を向上する試みが多く行なわれている。さらに、HMM-LR 連続音声認識方式などの手法では、音素継続時間を制御することができる。また、SSS-LR 連続音声認識方式では、音素や異音 (allophone) の継続時間を SSS アルゴリズムによりクラスタリングし、LR 構文解析の中で継続時間を制御している。

さらに音素以上については、日本語の場合は特にモーラ単位にも継続時間制約は存在する。このような、超分節的 (suprasegmental) な継続時間長 (duration) の制約 (constraint) ができれば、継続時間制約は音声認識に有効に利用できるだろう。しかし、その実現は、LR 構文解析でも、one-pass DP でも容易ではない。

そこで、行列パーザ (言語モデルの章を参照) の原理による、モーラタイミングの制約を含めたパーザを開発し、認識率の向上を図った [Singer93ICASSP04]。

6.2.4 ピッチ周波数パターンを用いた平叙文と疑問文の区別

これはまだ初歩的な検討で、まだ外部には発表していない。文末のピッチパターンの正規化と特徴量などを含めて検討した。

第 7 章

ニューラルネットワークによる音声認識の研究

7.1 TDNN を用いた音声認識

7.1.1 TDNN による音素認識

A. Waibel によって始められたニューラルネットワークを用いた音声情報処理の研究は Time-Delay Neural Network (TDNN) として定式化され、その有効性が示された。¹。従来から研究されてきた HMM に基づく音素認識系に比べて高い音素認識率を与えることを示したことは、多くの研究者を刺激し国内での音声認識に対するニューラルネットワーク研究の火をつけることになった。

[Waibel-87-11], [Waibel-87-12], [Waibel-88-04a], [Waibel-88-04b], [Waibel-88-05], [Waibel-88-08], [Waibel-88-10], [Waibel-88-11a], [Waibel-88-11b], [Haffner-88-10], [Haffner-88-11a], [Haffner-88-11b]。

7.1.2 TDNN-LR による連続音声認識

TDNN を用いた音素スポッティング法と拡張 LR パーザとを組合せ文節認識に適応し評価を行なった。これは基本的に HMM による音素認識と LR パーザとの組合せと同種のものである。前者はスポッティングを基本にしている点で動作が若干異なる。表 7.2 にその認識性能を示す。他の方法の音素認識率の比較を表 7.3 に示す。

[Minami-89-07], [Minami-89-10], [Minami-90-1], [Minami-90-2b], [Minami-89-10], [Miyatake-89-03a], [Miyatake-89-03b], [Miyatake-89-06], [Miyatake-89-08a], [Miyatake-89-10], [Sawai-89-03], [Sawai-89-05], [Sawai-89-07a], [Sawai-89-07b], [Sawai-89-08], [Minami-90-9a]。

参考のために評価に使用した音声資料の仕様を表 7.4 に示す。

¹分析条件などを以下の表 7.1 に示す。

表 7.1: TDNN 認識系における音声分析条件

発声者	男性 1 名 (MAU)
標準化周波数	12kHz
窓	256 点 Hamming 窓
更新周期	10ms
特徴量	16 帯域 FFT メルスペクトル
正規化	15 フレーム内で平均を 0 にした後、最大最小の絶対値の大きい方で正規化

表 7.2: TDNN-LR による文節認識率 (279 文節)

候補 順位	文節認識率 (%)		
	TDNN	HMM Single VQ	HMM Separate VQ
1	63.4	72.0	83.2
≤ 2	74.6	85.3	93.9
≤ 3	78.9	91.8	96.4
≤ 4	82.1	94.3	97.5
≤ 5	83.2	95.3	98.6

Single VQ: Power + WLR

Separate VQ: Power + WLR + DCEP

LR 探索 beam 幅 ≤ 18(local); ≤ 250(global)

表 7.3: 各種の方法による音素認識率 (18 子音)

認識方法	rank	単語発声	短い文節	自由発声
		5.7 モーラ	7.1 モーラ	9.6 モーラ
Discrete HMM	= 1	93.1	81.4	71.6
	≤ 3	99.5	95.9	92.4
Continuous HMM	= 1	98.1	79.7	66.6
	≤ 3	99.8	94.6	86.9
TDNN	= 1	96.2	76.2	56.6
	≤ 3	99.6	91.5	78.5
LVQ	= 1	97.9	81.7	61.6
	≤ 3	99.9	96.8	87.7
LVQ+HMM	= 1	97.2	80.6	69.6
	≤ 3	99.6	94.6	89.3
Fuzzy LVQ+HMM	= 1	94.4	80.8	74.0
	≤ 3	99.7	96.8	93.3

表 7.4: 音声資料の仕様

	発話速度 モーラ	記事
単語音声	5.68	5240 単語の奇数番目
短い文節	7.14	複合語を許さない文節発声
長い文節	7.72	文節発声
自由発声	9.56	区切り指定なしの文節発声

7.1.3 TDNN-LR による大語彙単語音声認識

TDNN の音声認識への 1 つの有望な応用として大語彙単語音声認識が検討されている [Minami-90-1]。辞書中の単語数を変化させた時の、特定話者に対する大語彙単語音声認識実験がなされ、単語数が増加するに連れて認識率は若干減少するものの、5240 単語辞書に対して 92.6% の高い認識率を実現している。文節音声認識においては複雑な調音結合のために単語音声で学習したニューラルネットワークの性能の劣化が起こる。同一の発話様式であるので大語彙単語音声認識はニューラルネットワークにおいては劣化が少ないものと思われる。

また、音声認識のためのニューラルネットワークの発火は音素に対する確からしさを与えることになる。この点に着目し、辞書中の音素テキストとニューラルネットワークの出力パターンに基づく単語予備選択が検討されている。大語彙単語音声認識で解決が必要である処理時間の高速化の検討がなされている [Fukuzawa93ASJ3]。

7.2 TDNN の改良 (問題点克服の試み)

その後の研究により、TDNN を用いた場合の認識性能が、HMM に基づく方法に比べて累積認識率という点で低くなっていることが明らかになった。この原因としては、ニューラルネットワークの出力値が 0-1 と離散的であり、2 位以下の認識スコアの持つ情報量が少ないためであると考えられる。一般的な HMM では、他のカテゴリの影響を考慮せず、もっぱらパターンの再生を主眼としてモデルが形成されるため、出力されるスコアには各カテゴリに対する尤もらしさ (尤度) の情報が含まれている。そこで、ニューラルネットワークを使用する場合でもこのような尤もらしさの情報を出力値に反映させることをめざして、幾つかの研究が行なわれた。

7.2.1 平滑化による頑健性の向上

ニューラルネットワークの出力や教師信号などの平滑化によって認識性能の向上をめざした研究が行なわれた。平滑化を行なう対象として、以下のようなものが提案されている。

- ニューロン出力の平滑化 [Minami-90-10]
- 学習における平滑化
 - － 連続値教師信号による学習
 - － 分布を平滑化した学習 [Kawabata-90-3]

以下でこれらの研究の概説を行なう。

出力における平滑化

[Minami-90-10] では、主成分分析 (5 次元) による直交基底ベクトルを算出しその方向の代表点に対する TDNN 出力値を用いて出力値を平滑化する方法が提案されている。TDNN を $f(x)$ と表す時、その平滑化関数 $\hat{f}(x)$ は以下のように定義される。

$$\hat{f}(x) = \frac{\int_{|x-y|<\epsilon} f(y) dy}{\int_{|x-y|<\epsilon} dy} \quad (7.1)$$

平滑化によって、通常の TDNN に比べて 2% 程度の識別率の向上が得られている。さらに、累積識別率においても通常の方法に比べて優れていることが示されている。積分する領域については、入力層において積分により平滑化するよりも第 1 中間層での平滑化の方が有効であることが示されている。主成分分析の次元の取り方、計算量の評価などの検討項目が残されている²。

KNIT 法

[Kawabata-90-3] では、KNIT 法と呼ばれる学習における平滑化手法が提案されている。従来のニューラルネットワークの学習が、入力空間から出力空間への点対応の学習であったのに対して、この方法では入力空間

² ϵ 近傍の取り方は、学習音声を認識するとき、認識率が最大になるように設定している。

表 7.5: KNIT 法を用いた母音認識実験結果

学習方法	学習 sample	認識率
従来法	closed	100.0
従来法	open	97.6
KNIT	closed	99.6
KNIT	open	99.2

の2点及びその対応する出力の2点を結ぶ線分の間に対応関係を学習させている。連続体(線分等)に対する学習を実現するためには連続体をいくつかの点で覆い尽くし、その点における対応関係を学習させる方法が考えられるが、そのためには学習時間が膨大になる。それに対して、無限小の近傍の2点に対する近似化した方式を用いることにより従来方法の数倍程度の計算量で、線分の学習を実現している。KNIT法を用いた母音認識結果を表7.5に示す。積分の持つ意味、積分区間を線分の中点としている点、離散量の教師信号に対する検討、等の検討項目が残されている。

ファジー (Fuzzy) 学習

[Komori91SP6]では、教師信号の平滑化が提案されている。この方法では、学習に用いる教師信号に確率のような連続値を与えることによって、尤もらしさを反映させたTDNNの学習をめざしている。連続的な教師信号は、学習パターンの近傍にある点の情報によって決定される。この方法によって識別率が向上することが報告されている。

[Komori91ASJ10], [KomoriIEICEtrans7], [Komori92ICASSP3]。

7.2.2 Network 構造の検討

[Minami-90-2b]では、TDNNの構造の面から認識性能の向上について検討されている。検討項目は以下の通りである。

- ネットワークモジュール構成の比較
 - 分散型
 - 集中型
- 出力、中間層間の接続法の比較
 - 自由な接続
 - 前部(4個)と後部(5個)とに分割
- 入力、中間層間の接続法の比較
- 出力、中間層間の重み係数に窓掛けする方法
- 周波数軸方向の重み係数の平滑化
- クラス分けネットワークの効果

これらの検討から以下の知見が得られている。

- クラス分けネットワークの有効性
- 入力層、中間層および出力層、中間層の間の shift invariant 接続の有効性

表 7.6: 対判定型 TDNN を用いた音素認識結果

(/b, d, g, m, n, N/)

発話様式	第1位	第2位	第3位
単語発声	95.2	99.2	99.8
	94.7	98.6	99.7
	97.3	99.2	99.8
文節発声	84.8	92.7	96.5
	80.6	93.9	98.1
	88.8	96.9	99.2
短い文節発声	84.5	91.2	95.3
	77.2	93.2	97.0
	86.2	95.4	98.1
自由発声	77.8	87.1	92.2
	73.2	90.2	96.3
	81.6	93.3	96.7

上段: 従来の TDNN

中段: 中間値学習なし PD-TDNN

下段: 中間値学習あり PD-TDNN

7.2.3 対判定型 TDNN(PD-TDNN)

TDNN を用いた音素識別において、第1位の識別率は非常に高いが、第2位以下の累積識別率はあまり向上されず、認識システムの中に組み込む場合の問題点となることが指摘されている。これを解決する方法として、従来の0-1学習によりカテゴリ間に形成されるcriticalな識別境界面の代わりに、0, 1, 0.5の3通りの教師信号を用いて学習された音素対間識別用TDNNに形成される緩やかな識別境界面を使用し、さらに認識時に多数の音素対間識別スコアの多数決によって最終候補を決定するという、対判定型TDNNを用いる音素認識手法が提案された。

[Takami-90-6][Takami-90-9][Takami-90-11][Takami91NNSP9]。

この方法では、音素 $/p_i, p_j/$ を判定する小規模なTDNNを作成し、 $/p_i/$ に対するTDNNの出力値が

$$\sum_{j \neq i} S(p_i | p_i : p_j) \quad (7.2)$$

のように対判定TDNNの出力の加算によって定義されている。ここで、 $S(p_i | p_i : p_j)$ は音素対 $/p_i, p_j/$ を識別するTDNNから得られる $/p_i/$ に対する出力であり、このTDNNは、入力音素が $/p_i/$ のときに1、 $/p_j/$ のときに0、それ以外の音素のときに0.5が出力されるように学習されている。さらに、このような中間値を教師信号として用いた独特な学習を円滑に進めるために、出力ユニットには2つのsigmoid関数を接続した形の非線形関数を用いている。

$$f(x) = \begin{cases} \frac{g(x + \alpha)}{2g(\alpha)} & (x < 0) \\ 1 - \frac{g(-x + \alpha)}{2g(\alpha)} & (x \geq 0) \end{cases} \quad (7.3)$$

この対判定型TDNNを用いた場合の $/b, d, g, m, n, N/$ に対する音素識別率を、比較のために行なった従来のTDNNおよび中間値学習なしの対判定型TDNNを用いた場合の音素識別率と合わせて表7.6に示す。その結果、中間値を教師信号とすることにより第1位の認識率、および第2位以下の累積認識率が共に改善されることが示されている。

表 7.7: 時間構造を考慮した TDNN による音素認識率

(/b, d, g, m, n, N/)			
入力音声	TDNN	TS	STS
短い文節	76.6	82.7	79.0
長い文節	75.9	77.6	77.7
自由発声	61.8	70.7	69.0
-20ms	91.2	52.7	58.7
-10ms	94.7	86.7	92.6
0ms	95.7	97.6	97.0
+10ms	94.1	84.4	89.9
+20ms	85.9	56.5	65.8

TS: Temporal Structure 型

STS: Shifted Temporal Structure 型

表 7.8: 時空間 Block 統合 NN を用いた音素認識率

(/b, d, g, m, n, N/)			
入力音声	TDNN	TF	BW
単語発声	95.9	96.7	98.2
短い文節	78.7	83.6	84.1
長い文節	79.8	84.3	83.9
自由発声	67.2	80.8	82.8

TF: Time Frequency TDNN

BW: Block Window NN

7.2.4 音声パターンの時間構造を考慮した TDNN

TDNN は音素識別性能が高く、シフトトレラントであることが示されているが、学習時と異なる発話様式の音声に対する識別能力はあまり高くない。音素の時間構造の特徴はより頑健であると考えられる。時間構造考慮 TDNN として、入力層に対する第1隠れ層（もしくはさらに第2隠れ層）に対して時間変化に対応する構造を導入する。音素の時間構造を中間層の4つの状態の連鎖で表現する。入力層と4つの第1隠れ層との結合は TDNN と同様である。ただし、そのシフトの範囲は15フレーム全体に渡るのではなくその部分に対応する。その他の結合は全結合とする。表 7.7 に述べるように実験の結果、連続発声の文節中の音素識別 (/b, d, g, m, n, N/) に対して有効であることがわかった。ただし、シフトトレランスに対しては robust でなくなる問題点も指摘されている [Komori-90-3b]。

7.2.5 時空間 Block を統合したニューラルネットワーク

TDNN はその構造に shift invariant (重み係数の tied connection) を導入することにより、時間方向への音素パターンのずれに対する耐性を高めている。音声における音素パターンの変動は時間方向へのものと周波数方向へのものがあり、これが認識を難しくさせるおおきな原因となっている。そこで、この shift invariant 構造を周波数軸にも拡張しようとする研究が行なわれた。[Sawai-90-9b] では2つの方法が提案されている。第一の方法では、周波数軸におけるずれを吸収する構造を導入し、時間軸でのずれを吸収する従来の TDNN と上位の層において統合する構造 (TF) が考えられている。また第2の方法は、時空間パターンの1つのブロックを上位の層に統合する構造 (BW) である。入力層、第1、第2隠れ層において、時間×周波数を 3×4 , 5×5 , $5 \times$

5ごとに上位の層に統合させている。TFにおいてはTDNNと同様に tied connection を周波数軸方向にも入れている。また一方、BWにおいては重み係数はすべて free connection としている。これら2つの構造と従来のTDNNとの性能の比較を表7.8に示す。この結果から上の2つの構造は単語発声においても有効であるばかりでなく、異なる発話様式の音声に対しても顕著な認識率の改善を実現している。

7.3 確率的出力をもつニューラルネットワーク (Fuzzy Partition Model)

TDNNの音素スキミングパターンから全ての音素クラスが無発火であったり、複数の音素クラスが同時に大きく発火するという現象が確認されていた。この現象は連続音声認識の性能を劣化させる原因であり、それを回避するために確率的な出力をもつニューラルネットワークである Fuzzy Partition Model と呼ばれるアーキテクチャを導入した。

FPMは多入出力素子で構成されるニューラルネットワークであり、従来のパーセプトロンタイプのネットワークとは異なるアーキテクチャを持つ。第 m 層第 s 素子の中の k 番目の入力、出力をそれぞれ $u_{km}^{(s)}$, $a_{km}^{(s)}$ とし、 $a_{jm-1}^{(g)}$ と $u_{km}^{(s)}$ とを介する重みを $w_{kmjm-1}^{(sg)}$ とおけば素子の入出力関係は次式で表すことができる。但し、第 m 層の素子数を M^m と表す。

$$a_{km}^{(s)} = \frac{\exp(u_{km}^{(s)})}{\sum_{j=1}^{N^m} \exp(u_{jm}^{(s)})} \quad (k=1, \dots, N) \quad (7.4)$$

$$u_{km}^{(s)} = \sum_{g=1}^{M^{m-1}} \sum_{j=1}^{N^{m-1}} w_{kmjm-1}^{(sg)} a_{jm-1}^{(g)} \quad (k=1, \dots, N) \quad (7.5)$$

式(7.4)は、 $\exp(u_{km}^{(s)})$ を正規化することを意味するから、素子内の出力は常に正でその総和は1となり、1素子内の出力群の関係は次式で制限される。

$$\sum_{j=1}^{N^m} a_{jm}^{(s)} = 1 \quad (7.6)$$

$$0 \leq a_{jm}^{(s)} \leq 1 \quad (\forall j) \quad (7.7)$$

次にFPMの学習アルゴリズムについて述べる。学習には最急降下法を用い、出力層の教師と出力との誤差が最小になるように素子間の重みを逐次的に変化させてモデルを収束させる。誤差評価関数にはFPMの出力が確率的であることからKullbackダイバージェンス D を採用している。Kullbackダイバージェンスを用いることにより高速な学習を可能にする。学習率を η とすれば重み修正量は次式で与えられる。

$$\Delta w_{kmjm-1}^{(sv)} = -\eta \frac{\partial D}{\partial w_{kmjm-1}^{(sv)}} \quad (7.8)$$

$$D = \sum_{g=1}^{M^m} \sum_{j=1}^{N^m} t_j^{(g)} \log \frac{t_j^{(g)}}{a_{jm}^{(g)}} \quad (7.9)$$

式(7.9)を式(7.8)に代入し、過去に変更した修正量の影響量 α も考慮すれば次式になる。

$$\Delta w_{kmjm-1}^{(sv)} = \eta \delta_{km}^{(s)} a_{jm-1}^{(v)} + \alpha \Delta w_{kmjm-1}^{(sv)} \quad (7.10)$$

$\delta_{km}^{(s)}$ は層によって異なり、 m が出力層の場合には式(7.11)、それ以外の場合には式(7.12)となる。

表 7.9: 6 子音認識結果 (単位 %)

モデル	単語発声	文節発声	短い文節発声	文発声
TDNN	95.2(99.8)	84.8(96.5)	84.5(95.3)	77.8(92.2)
FPM	94.0(99.9)	79.3(97.8)	80.3(97.4)	76.1(96.5)

(括弧内は第3位までの累積認識率)

表 7.10: 18 子音認識結果 (単位 %)

モデル	単語発声	文節発声	短い文節発声	文発声
TDNN	92.0(99.3)	73.7(89.2)	—	62.4(78.5)
FPM	90.5(99.4)	76.5(95.8)	76.5(95.1)	67.6(91.0)

(括弧内は第3位までの累積認識率)

表 7.11: 278 文節音声認識結果 (単位 %)

モデル	認識率
TDNN	66.9(82.4)
FPM	72.7(90.3)

(括弧内は第3位までの累積認識率)

$$\delta_{k^m}^{(s)} = t_{k^m}^{(s)} - a_{k^m}^{(s)} \quad (7.11)$$

$$\delta_{k^m}^{(s)} = a_{k^m}^{(s)} (\sigma_{k^m}^{(s)} - \sum_{j=1}^{N^m} a_{j^m}^{(s)} \sigma_{j^m}^{(s)}) \quad (7.12)$$

但し、

$$\sigma_{k^m}^{(s)} = \sum_{g=1}^{M^{m+1}} \sum_{i=1}^{N^{m+1}} w_{i^{m+1}k^m}^{(gs)} \delta_{i^{m+1}}^{(g)} \quad (7.13)$$

表 7.9, 及び 7.10 に 8 音素及び 18 音素認識実験の結果を示す。また、表 7.11 に 278 文節認識実験の結果を示す。実験結果から発話様式の異なるセットに対して TDNN よりも高い認識率が得られ、FPM によるアーキテクチャが発話様式に対して頑健であることを示した。

[Kato92ASJ03][Kato92IEICE06][Kato92NNSP08][Kato92ASJ11]

7.4 不特定話者音声認識

ニューラルネットワークを用いて不特定話者音声認識を実現するための方法として、不特定対応大規模ネットワークを用いる方法 [Sawai91ASJ3][Nakamura91ASJ10][Nakamura92ICASSP3] と話者適応を用いる方法とが提案された。不特定話者認識 TDNN は大規模となるためその学習が容易ではない。それに対して、話者適応は少量の音声を用いて学習ができる点で現実的な方法である。時間構造を持った音響セグメント間のニューラルネットワークによる話者適応の有効性が示されている。

[Fukuzawa-90-9][Fukuzawa91ASJ10a][fukuzawa91NNSP9][Fukuzawa92ICASSP3]

7.5 識別誤り最小規準による学習

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 [MacDermott-SP91-12]。[Sugiyama92ASJ10] 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検討している。

7.6 堅牢なシステムの構築に向けて

ニューラルネットワークを用いたより高性能なシステムの構築に向けて、幾つかの試みがなされている。新たな連続音声認識向けのアーキテクチャーの導入と文節発音音声を用いた追加学習法を用いて、高い認識率が実現されている [Kato92IEICE02] [Kato92ASJ3]。また、雑音環境下での堅牢性の評価と雑音対策について検討されている [Ohkura91ASJ3]。

7.6.1 FPM-LR による不特定話者連続音声認識

ニューラルネットワーク (NN) 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、不特定話者の音素パターンを 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、学習に膨大な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。

[Fukuzawa92ASJ10][Fukuzawa92SP12][Fukuzawa92ICSLP10]

7.6.2 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取出すことができる。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。

[Kato92ASJ10][Ioana93SP1][Kato93ICASSP4]

言語処理への応用

音声処理などの言語処理のために数理的な言語モデルが研究されており、文脈自由文法 (CFG) およびその自動学習、確率付きモデルのパラメータ推定、多重マルコフモデル (N-gram) を用いた研究などがなされている。ニューラルネットワークを用いた言語処理への応用として単語の品詞予測を行なう Net-gram に関する検討がなされている。

[M.Nakamura-88-03][M.Nakamura-88-06][M.Nakamura-88-10] [M.Nakamura-88-12][Nakamura.M-90-08] 実験結果から従来の N-gram と同程度の性能であることが示され、ニューラルネットワークの応用範囲を拡大している。

7.7 離散値写像ニューラルネットワークと連続値写像ニューラルネットワーク

ここまで述べたニューラルネットワークは、音素カテゴリーを認識する、いわゆる識別型のものであった。それに対して、ニューラルネットワークの持つ非線形性を利用した非線形写像による音声処理の研究が行なわれている。雑音抑圧への応用、恒等写像を用いた音素認識、話者適応について述べる。

7.7.1 波形上での雑音抑圧

雑音抑圧への応用は、ニューラルネットワークによる非線形写像の1つの例とし初期の頃から研究された。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。[Tamura-88-01][Tamura-88-03][Tamura-88-04] [Tamura-88-10][Tamura-89-01][Tamura-89-05]

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。さらに、表7.12に示すように 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が得られている [Ohkura-90-9]。

表 7.12: 各種の雑音抑圧手法を用いた音素認識の比較

SNR	入力	評価方法			
		無処理	NN	CM	NN+CM
5 (dB)	M1	43.8	59.9	58.3	62.3
	M2	35.5	50.9	63.6	57.2
	M3	33.7	64.4	61.5	59.6
	M4	33.9	52.7	60.4	55.2
20 (dB)	M1	61.0	59.9	68.9	62.1
	M2	62.3	55.1	56.2	61.1
	M3	56.3	64.3	72.7	63.9
	M4	53.1	60.2	61.7	61.3

NN: 雑音抑圧 NN を使用

CM: コードブックマッピングを使用

M1: WLR+DCEP+POW, M2:WLR+DCEP

M3:WLR+POW, M4:WLR

7.7.2 音素ニューラルフィルター (PFN)

各カテゴリ毎に多層ニューラルネットワークを用いて、恒等写像を構成することにより、そのカテゴリ特徴をネットワークに学習させるものである。TDNN による認識においては第一位での認識率は高いが第2位以下の累積認識率が他の認識モデルに比べて低いという問題点が指摘され、その解決方法として、提案されたものである。

[Nakamura-90-3][Nakamura-90-6][Nakamura-90-9b][Nakamura-90-11][Nakamura91ICASSP5]

入力音声 X は音素カテゴリ毎に作成された NN に入力され、その出力との類似度が計算される。

$$S = \frac{(V_{in}, V_{out})}{\|V_{in}\| \|V_{out}\|} \quad (7.14)$$

この類似度を用いて、入力音素のカテゴリを判定する。

この構造は、磯による NN 予測問題 [Iso-SP89-23] と関連が深く、過去の点を用いない、即ち縮退した場合と見ることもできる。ニューラルネットワークの新たな可能性を追求しているという点で注目される方法であるが、一方では NN の持つ識別率を最大にする学習という枠から外れてしまう点で、性能の劣化が懸念される。

7.7.3 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。[S.Nakamura-89-02][S.Nakamura-89-03][S.Nakamura-89-08f] 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、その有効性を示している。

[Fukuzawa-90-9][Fukuzawa91ASJ10a][fukuzawa91NNSP9][Fukuzawa92ICASSP3]

音素認識結果を表 7.13 に示す。

表 7.13: セグメント話者適応 NN を用いた音素認識

	(/b, d, g/)	
	未知話者	標準話者
適応前	83.6	98.7
適応後 1	86.6	-
適応後 2	67.5	-

適応後 1: 恒等写像 NN を用いた話者適応

適応後 2: ランダム weight 値からの話者適応

未知話者:MHT, 標準話者:MAU

7.7.4 教師なし構造変換ニューラルネットワーク

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、教師なし話者適応において成功を収めてきた教師信号を自動的に生成し写像を構成する方法に比べて、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 [Sugiyama-90-9]。音素認識実験を通してその有効性が示されている。[Fukuzawa91ASJ10b][Fukuzawa92IEICE9]

第 8 章

知識処理による音声認識の研究

8.1 知識処理に基づく音声認識

8.1.1 スペクトログラムリーディング知識に基づく音声認識

人間のスペクトログラムリーディング知識を、規則の形で機械に教えることにより、音声認識をするエキスパートシステムを作ることができる。ATR 発足当時は、このような考え方が有望視されていて、活発に研究が行なわれた。

文献: [Hatazaki-87-9][Kawabata-87-10][Hatazaki-88-01] [Hatazaki-88-03][Hatazaki-88-08][Hatazaki-88-10] [Hatazaki-89-03a][Hatazaki-89-03b][Hatazaki-89-03c] [Hatazaki-89-05][Komori-89-03] など。

8.1.2 知識とニューラルネットに基づく音声認識

上記の、知識に基づく手法は、音素のセグメンテーションには向くが、音素識別性能は余り高くできなかった。一方、ニューラルネットワークの研究が進むにつれて、ニューラルネットワークを用いれば高い音素識別性能が得られることがわかってきた。そこで、知識とニューラルネットを統合した音声認識エキスパートシステムについて研究し、より高い認識性能を得るために、知識とニューラルネットの統合方法やそれぞれの音声認識技術の拡張・改良を行った [Komori-89-06][Komori-89-09][Komori-89-10]。

音素認識エキスパートシステム

知識とニューラルネットを統合した子音認識エキスパートシステムを構築した。音素認識実験の結果、整合性を考慮した統合方法は子音認識率 89.4% を示し、また、セグメンテーション性能も向上することも示した。更に、挿入誤り率が 27.8% から 12.4% に大幅に削減できることも示し、提案するシステムの有効性を示した。

知識とニューラルネットを用いた母音認識部を統合し、全音素の認識を行う音声認識システムを構築した。単語辞書等の言語情報を全く用いないで全音素認識を行った結果、音素認識率で 91.4%、脱落誤り率 3.6%、置換誤り率 5.0%、挿入誤り率 20.7% が得られた。

音声認識エキスパートシステムの連続音声・複数話者音声への拡張

提案した音声認識システムの音素セグメンテーション部の連続音声や複数話者音声への耐性およびその拡張可能性の検討を行った。少数でかつ簡単な知識の変更により複数話者音声への耐性を強化でき、比較的良好な性能を示すことが実験より判明した。特定話者文節音声に対して音素セグメンテーション率 88.4%、平均境界誤差 5.5ms、複数話者単語音声に対しては、91.1%、6.2ms であった。特定話者単語音声に対しては、93.3%、5.8ms である。

Time-State Neural Networks (TSNN)

音素識別用ニューラルネットの発話速度変動に対する耐性を強化する構造を持つ「Time-State Neural Networks」(TSNN)を提案した。音素には、発話速度等が変化しても大きく変わらない音素特有の時間構造があり、この時間構造を音素識別用ニューラルネットで扱えば、発話速度等に耐性のあるニューラルネットを構築することができる。この時間構造を扱える TSNN を提案した。/b, d, g, m, n, N/ の音素識別実験の結果、従来法 TDNN では、音素識別率は単語発声 95.7%、文節発声 76.8%、文発声 61.8% が、TSNN では、97.0%、78.8%、71.1% に向上した。

ニューラル・ファジー学習法

音素識別用ニューラルネットの発話速度変動に対する耐性を強化する学習方法「ニューラル・ファジー学習法」を提案した。提案学習方法では、0と1の2値による確定したクラス情報を教師信号とする従来学習法と異なり、非確定的(ファジー)なクラス情報教師信号とする。本方法により、過学習を回避し、汎化能力や累積音素認識率が向上し、耐性のあるニューラル・ネットの構築を目指す。音素識別実験では認識率が向上し、特に TDNN-LR 連続音声認識システムを用いた文節音声認識実験では、第1位認識率、従来学習法 71.2% が提案学習方法 80.9% に、累積第5位認識率 92.8% が 96.0% に向上した。更に、提案学習法が高速な学習方法であることも実験より判明した。

8.1.3 スペクトログラムリーディング知識に基づく自動ラベリング

自動翻訳電話を実現するための基盤技術の研究として、音声認識および音声合成の研究を精力的に進めた。音声認識・合成など、いわゆる音声情報処理システムの研究開発には、整備された大規模な音声データベースが必要であり、特に、音素の位置情報を表す音素ラベルデータは重要な役割を果たす。例えば、認識システムでは音素モデル(HMM, neural network 等)のパラメータの推定(学習)に用いられ、合成システムでは合成用音素片データの作成に利用される。

音素ラベルデータは、スペクトログラムの視察という手作業により作成されるため、大量の音声ラベルデータの作成には膨大な時間・コストを費し、また複数の作業員間でのデータ統一性を保つための努力が必要である。従来の音素ラベリング作業におけるこのような問題点を克服するため、音素ラベリング作業を自動化するための手法を研究した。さらに、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として、実用的な音素ラベリングシステムを開発した。

HMM と知識に基づく音素セグメンテーションシステム

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 [Fujiwara91ASJ3]。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 [Fujiwara91KOREA7, Fujiwara91ASJ10, Fujiwara92ISSPA92]。

音素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 [Fujiwara92ICSLP10, Fujiwara92ASJ10, Fujiwara93IEICE1]。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。

第 9 章

諸方式の音声認識性能比較

9.1 連続音声認識システム群 ATREUS

自動翻訳電話のための連続音声認識の研究の結果、多数の連続音声認識システムを構築した。このシステム群を ATREUS と名付けた。これらの同一のタスクや条件での比較を、以下に示す。完全に同じ条件ということは、別別の方式の場合はほとんど不可能であるので、このような比較には限界があることを了承して頂きたい。

ATREUS 音声認識システム群の実験タスクは、「国際会議参加問い合わせタスク」である。

表のスペースの制約から、システムの構成要素に以下のような記号をつける。

Acoustic analysis

- (A1) Fuzzy Vector Quantization
- (A2) LPC analysis
- (A3) Mel-scaled FFT-based spectra

Acoustic models (1) — Hidden Markov Models

- (M1) Discrete density HMM based on Fuzzy Vector Quantization and multiple code-books (FVQHMM)
- (M2) Continuous Mixture Output Probability Density HMM (CMHMM)
- (M3) Hidden Markov Network (HMnet)

Acoustic models (2) — neural networks

- (M4) Time-delay Neural Network (TDNN)
- (M5) Pairwise-Discriminant TDNN (PD-TDNN)
- (M6) Fuzzy Partition Model (FPM)

Speaker-adaptive speech recognition

- (SA1) Codebook Mapping (CBM)
- (SA2) Vector Field Smoothing (VFS) for discrete HMMs
- (SA3) Vector Field Smoothing (VFS) for continuous HMMs
- (SA4) Speaker-tied training
- (SA5) Segmental speaker mapping using neural network

Speaker-independent speech recognition

(SI1) Fuzzy vector quantization HMM

(SI2) Continuous mixture HMM

(SI3) Speaker-mixture HMnet

(SI4) Speaker independent TDNN

(SI5) Speaker independent FPM

Parsing algorithms

(P1) Generalized LR Parser

(P2) One-pass DP [Murakami92]

(P3) Matrix parsing algorithm [Singer92e]

Grammars

(L1) Context Free Grammar (CFG) [Hosaka92]

(L2) Word bigram/trigram [Murakami92]

9.1.1 特定話者文節音声認識の性能比較

以下に特定話者文節音声認識の性能比較を示す。

表 9.1: Phrase recognition rates (%) speaker dependent systems

system	FVQHMM-LR		CMHMM-LR		SSS-LR		TDNN-LR		PD-TDNNM-LR	LR	
acoustic analysis	A1		A2				A3				
acoustic model	M1		M2		M3		M4		M5	M6	
training data	word		word+ phrase	word	word+ phrase	word	word ^{†2}	word			
language model	L1										
vocabulary size	≈ 1,000 words										
parsing algorithm	P1										
beam width	256		600		256		100				
duration control ^{†1}	no	yes	yes	no	no		yes				
#speakers for evaluation	1	3	1	3	1	1	1	2	2	1	4
testing speech data	279 phrase utterances										
rank 1st	75.6	89.8	88.9	91.3	93.2	93.2	95.3	65.6	71.2	75.5	76.3
≤ 5th	94.3	99.4	99.6	99.3	99.3	99.6	100.0	88.2	92.3	92.5	91.6

^{†1} state duration control for HMMs and phoneme duration control for TDNNs and FPM^{†2} Neural Fuzzy training

9.1.2 話者適応文節音声認識の性能比較

以下に話者適応文節音声認識の性能比較を示す。

表 9.2: Phrase speech recognition rates (%) in speaker adaptive systems

system	FVQHMM-LR					CMHMM-LR			SSS-LR				TDNN-LR
acoustic analysis	A1					A2							A3
acoustic model	M1					M2			M3				M4
speaker adaptation	SA1		SA2			SA3				SA4		SA5	
#reference speakers	2 male					1 male			1 male		3 male		2 male
#words for speaker adaptation	25	100	5	10	25	50	100	102 phrases	100	1	5	100	
language model	L1												
vocabulary size	≈ 1,000 words												
parser	P1												
beam width	256					600			256				100
duration control†	yes					no							yes
#input speakers	2 (1 male + 1 female)					2 (1 male + 1 female)			3 (2 male + 1 female)		3 male		2 (1 male + 1 female)
testing speech data	279 phrase utterances												
rank 1st	71.0	74.2	63.3	71.0	73.5	71.5	81.7	85.1	79.8	89.7	79.9	80.5	57.4
≤ 5th	93.5	95.2	88.2	92.7	93.9	89.8	93.5	95.2	96.2	99.4	97.2	97.2	78.5

†: with state duration control for HMMs and phoneme duration control for TDNN

9.1.3 不特定話者文節音声認識の性能比較

以下に不特定話者文節音声認識の性能比較を示す。

表 9.3: Phrase speech recognition rates (%) of speaker independent systems

system	FVQHMM-LR	RCMHMM-LR	SSS-LR	TDNN-LR	FPM-LR (multi)	
acoustic analysis	A1	A2		A3		
acoustic model	M1	M2	M3	M4	M6	
speaker model	SI1	SI2	SI3	SI4	SI5	
training speech data	words (200P ^{†2})	words (216B ^{†3} +520)	words (216B)	words (2620)	words (2620+216B)	words + phrases
#speakers for training	10 male	12 male		8 male	16 (8 male + 8 female)	
language model	L1					
parsing algorithm	P1					
duration control ^{†1}	yes	no		yes		
beam width	256	1200	256	100		
#testing speakers	10 male	10 male		2 male	4 (2 male + 2 female)	
testing speech data	279 phrase utterances					
rank 1st	74.5	75.2	82.6	68.3	70.0	80.0
≤ 5th	95.1	91.8	98.6	89.6	88.5	93.6

^{†1} state duration control for HMMs and phoneme duration control for TDNN and FPM

^{†2} 200 phonemes

^{†3} 216 phonetically balanced words

9.1.4 特定話者文音声認識の性能比較

以下に特定話者文音声認識の性能比較を示す。

表 9.4: Sentence speech recognition rates (%) in speaker dependent systems

system	FVQHMM-LR		SSS-LR	HMM/bigram	HMM/trigram
acoustic analysis	A1		A2	A2	
acoustic model	M1		M3	M2	
training data	word + phrase + sentence			word	
language model	L1			L2	
grammar	CFG only	stochastic CFG, rule bigram	CFG only	word bigram	word trigram
parsing algorithm	P1			P2	
duration control†	yes		no	no	
beam width	100			4096	
testing data	137 sentence utterances			38 sentences	
vocabulary	744			1500	
#speakers	1 male			1 male	
testing speech data	137 sentences (continuously uttered)				
rank 1st	57.7	83.9	64.2	57.9	76.3
≤ 5th	65.5	86.1	66.4	-	-

†: with state duration control for HMMs

9.1.5 1500 語、3000 語の語彙の文節音声認識の性能比較

以下に、1500 語、3000 語の語彙の場合の文節音声認識の性能比較を示す。

表 9.5: Phrase speech recognition rates in 1500- and 3000-word vocabulary speaker-adaptive/dependent SSS-LR

system	SSS-LR		
acoustic analysis	A2		
acoustic model	M3		
speaker adaptation algorithm	SA3 (VFS)†	speaker-dependent	SA3 (VFS)†
#reference speakers	6 (3 male + 3 female)	1 male	6 (3 male + 3 female)
#words for speaker adaptation	25	none	25
language model	L1		
vocabulary	1500 words	3000 words	
parser	P1		
duration control‡	no		
beam width	250		
#input speakers	4 (2 male + 2 female)	1 male	4 (2 male + 2 female)
testing speech data	703 phrase utterances		
rank 1st	90.2	89.8	87.2
≤ 5th	97.7	95.6	95.4

†: Before speaker adaptation, the acoustically closest reference speaker to the input speaker is pre-selected automatically.

‡: with state duration control for HMMs

第 10 章

音声分析、言語識別、音声区分化、その他の研究

10.1 音声認識のための音声分析の研究

10.1.1 ベクトル時系列のフィルタによる特徴量抽出

音声認識において広く用いられている音声の動的な特徴量である、いわゆるデルタケプストラムは、当該フレームを中心として、重みつき最小二乗法によって、線形近似を行ない、その傾斜の係数ベクトルを動的特徴量として用いるものである。

その計算式を検討すると、ケプストラムベクトル系列とあるスカラー値系列の畳み込みになっていることが分かる。従って、デルタケプストラムとは、ケプストラムベクトル系列をある種のデジタルフィルタに通したものに他ならない。この考え方を拡張すれば、複数のデジタルフィルタをうまく設計すれば、今以上に有効な音声認識パラメータが得られるかも知れない。

このような観点から、いくつかの型のデジタルフィルタの設計と、それを用いた音素認識実験による評価を行なった [Katagishi92ASJ10][Katagishi92SST12]。

10.1.2 Dynamic cepstrum の検討

上記のデジタルフィルタは、スカラー係数であったが、これをマトリクス係数にすると、さらに広い音声特徴量のクラスが考えられる。その中に、相川らが提案した dynamic cepstrum があり、音素認識実験によってその有効性を評価した [Singer92ASJ10b][Katagishi92SST12]。

10.2 言語識別の研究

10.2.1 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 ケ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。本報告では、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。

- 入力単語数が増えるに従って、識別率は上昇する。どちらの尺度を用いてもある程度の大きさの符号帳 (16, 32 符号) を用い 4, 5 単語入力すれば日英の区別はほとんどできる。
- 符号帳が 8, 16, 32 の場合には入力単語数 (4, 5 単語) がある程度あれば非常に高い識別率である。
- 符号帳の大きさが 4 の場合には入力単語数を増加してもあまり高い識別率を得られない。
- WLR は CEP に比べて言語による歪みの大きさの変動が小さい。

10.2.2 スペクトル遷移の確率モデルによる言語識別

筆者等は、静的な特徴を用いた bilingual 話者音声の日英言語識別の有効性を報告した。本報告では同一タスクに対し、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。

10.3 音声の区分化・識別問題の研究

10.3.1 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。本報告では、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する。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 符号帳の符号の出現確率分布の N 個のカテゴリへのクラスタリングによる方法を提案する。セグメンテーションが未知の場合については HMM の適用が可能である。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。Universal VQ 符号帳の符号の出現確率分布のクラスタリングによる方法を提案し、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。[SP1992-07][SP1992-10][ICASSP93a][IEICE93-N-source][IEICE1992-09][TR-I-0262]

10.3.2 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。本報告は、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する。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。[ASJ1993-3-Sugiyama]

10.4 その他の研究

10.4.1 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る。[SP1990-05][ICSLP1990][IEICE1991-03][TR-I-0221]

10.4.2 伝送誤りに強い VQ 符号帳構成法に関する検討

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。[TR-I-0223]

10.4.3 LPC ケプストラム係数の存在領域について

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 Δ CEP の性質についても述べた。[ASJ1991-03a][SP1992-02] [IEICE93-Cep]

第 11 章

音声の規則による合成の研究

我々は自然で高品質な規則合成音声作成のため、複合音声単位を用いた音声合成方法、及び統計的手法を用いた韻律制御の最適化について検討を行なった。これらの検討結果に基づき ATR ν Talk 合成システムを作成した。ATR ν Talk 合成システムは、規則による音声合成の方法を提供すると共に、音声合成の工学的諸問題を、限られた研究者のノウハウとしてではなく、明確な数理的なアルゴリズムの問題として取り扱えるパラダイムを提供することを意図している。ATR ν -Talk システムの特徴は、次の二点にある。

- 非均一な構造を持つ合成単位 (Non-Uniform Unit) の選択的利用
- 韻律制御規則の最適化 (Automatic Tuning of Rules for prosody control)

11.1 非均一合成単位を用いた音声合成方式

文献 [Sagisaka-87-10],[Sagisaka-88-03] では、合成単位の種類、個数、抽出環境を固定せず種々の非均一音声単位を備え、使用環境に柔軟に対処した単位選択を行なう音声合成方式を提案した。文献 [Takeda-89-10],[Abe,K89-11] では、知識主導型の音声単位選択方法と選択された音声単位の単位境界近傍の音声環境に応じた接続方法を提案し、その有効性を確認した。さらに、文献 [Iwahashi91ASJ3],[iwahashi91SP5],[iwahashi92ICASSP3] では、知識主導型の音声単位選択方法で生じるシステムのチューニングの難しさを回避するために、音響的尺度に基づく単位選択法を提案した。この方式では、合成単位の抽出された音韻環境の適切さや、音声単位間の結合の連続性などが音響的に評価される。さらに、この音響的評価規準に基づいた最適な合成単位系列が、効率的な組合せ最適化手法により、合成単位データベースから選択される。

11.2 統計的手法を用いた韻律制御の最適化

(1) 基本周波数の制御

文献 [Sagisaka-86-12-b] では、付属語、複合語に関するアクセント規則化について述べた。文献 [Sagisaka-90-1],[Sagisaka-90-9a] では、統語構造に基づく日本語イントネーションの制御を目指し基本周波数パターン概形の制御の定量化を試み、その有効性を示した。また、統語境界におけるアクセント句全体の基本周波数の上昇現象を、境界直前の句が関与する範囲の統語構造の観点から分析を行なった。さらに、文献 [kaiki92IEICE05] では、規則による合成音声の品質向上のため、適切な基本周波数制御の規則化を目指して、局所的句構造に基づき基本周波数の定量的な分析を行なった。分析によって明らかになった現象を、基本周波数のパターン概形の制御に取り入れることを試みた。4 話者、各約 500 文の音声データベースを用いて、各アクセント句の先頭値・最大値・最終値の 3 点の基本周波数推定を行なった。従来の推定結果に対して、学習データで 20.3%、未知データで 14.6% (基本周波数対数値の平均二乗誤差) の推定精度向上を図ることができ、新たに用いた基本周波数推定モデルの有用性が確認された。

(2) 音韻継続時間の制御

文献 [Takeda-88-03],[kaiki-90-5],[kaiki-90-9a] では、音韻継続時間を制御する要因について、多量の自然音声データを基に統計的手法を用いて規則の定量化を図った。さらに、文献 [Kaiki91ASJ10a][Kaiki91ASJ10b] では、音韻継続時間の個人性の分析、文音声におけるポーズ長制御の検討について述べている。また、文献 [Kaiki92IEICE03] では、自然な文音声規則により合成するため、母音継続時間の制御の精密化を図った。これまで日本語の母音継続時間について指摘されてきた制御要因に加え、新たに、句構造や語彙等の言語情報が制御に及ぼす影響を調べるため、各母音に対して句内位置と形態素情報(品詞)を要因に加えて重回帰分析(数量化I類)を行なった。この結果、従来主として単語音声の分析で指摘されてきた要因の文中における影響が定量的に明らかになった。また、これら制御要因による推定精度を調べた結果、学習データで平均二乗誤差 15.30ms(標準偏差の 52.7%)、未知データで平均二乗誤差 15.84ms(標準偏差の 59.2%)となり規則の妥当性を確認できた。

(3) パワーの制御

これまで、基本周波数、音韻継続時間については、種々の手法を用いて定量的な分析及び規則化が行われてきた。しかし、音声パワーについては、あまり定量的な分析はされていなかった。また、規則合成システムにおいて単語中から音声単位を切り出して用いる場合、従来はその音声単位のパワー時間変化パターンを用いて適当に結合する方法がとられていた。しかし、単語中の素片のパワーを文章中のパワーとしてそのまま適用できるかは自明ではなかった。そこで、文献 [Abe,K-90-3],[Mimura91SP5],[Mimura92ASJ3] では、このようなパワー制御の問題について検討した。統計的手法を用いた母音区間の平均対数パワー(母音パワー)と合成に用いる音声単位の音声パワー時間変化パターンについて分析を行なった。この結果、母音部の平均対数パワーを制御する要因には、基本周波数、隣接音韻の種類、文・呼吸段落内位置、当該音韻の種類がありことが判った。また、それらの影響を文章及び単語音声データにより定量的に確認した。また、音声パワー時間変化パターンについて分析した結果、単語中のパワーパターンを大きく変えずに文章中のパワーパターンを実現できることが判った。さらに、分析合成音を用いた聴取実験により、制御モデルの有効性を確認した。

(4) 合成単位セットの最適構成法

文献 [Iwahashi92ASJ3],[Iwahashi92SP9],[Iwahashi92ASJ10],[Iwahashi92ICSLP10] では、高品質な音声合成される合成単位セットを、所望のデータ容量で構成する方法として、音声素片ネットワーク(SSN:Speech Segment Network)最適化法を提案する。本手法は、音声素片により構成されたネットワークに対してコスト関数を定義し、反復改善法やシミュレーテッド・アニーリング法などの組合せ最適化法を用い、大容量データベースから、素片歪みと接続歪みを同時に最小化する音声合成単位セットを選択するものである。DIPHONE 単位セットの構成実験を行なった結果、本手法を用いることにより接続歪みの総和で約 35%、最大値で約 70%の低減が確認され、この効果は素片候補数が増すほど大きくなることが示された。

(5) 種々の発話様式における韻律パラメータの性質

文献 [Miyatake-87-10b] では、多様な口調を持つ音声の合成を目指して、種々の発話様式で発声した単語及び会話文章データを分析した。この結果、韻律パラメータ相互間には依存関係があり、特にパワーと基本周波数の間には強い正の相関関係が見られることを示した。また、種々の口調における韻律の特徴を示した。

11.3 ATR ν Talk 合成システム

文献 [kaiki-TR-I-0285] では、ATR ν Talk 合成システムの構成について述べられている。また、[Mimura92ASJ10] では、なじみ度合の違う単語セットを用いた明瞭度試験を行ない、ATR ν -Talk 合成音声の単語明瞭度について分析した。音節明瞭度は、なじみのある単語では自然音声とほぼ同じであった。しかし、単語のなじみ度合が低くなるほど、自然音声との差が大きくなっていった。また、CV 単位を音声単位として作成した合成音声と比べた分析より、ATR ν -Talk で用いている接続歪みを最小化する音声単位選択法の有効性を確認した。

第 12 章

音声変換 (話者変換) の研究

12.1 声質変換

ある話者の声質を別の話者の声質に変換する声質変換の技術は、自動翻訳電話実現のために、重要な技術である。この技術に関して、まず、コードブックマッピングに基づいたアルゴリズムが開発された。その後、この技術を発展させ、異言語間の声質変換、およびセグメントを単位とする声質変換に関する研究が行なわれた。声質変換に関しては、以下の研究を行なった。

- ・ コードブックマッピングによる声質変換
- ・ 異言語間の声質変換
- ・ セグメントを単位とする声質変換

12.1.1 コードブックマッピングによる声質変換

ベクトル量子化された音声のマッピングによる声質変換法を提案する。本手法は、2人の話者に同一単語を発声させ、これらの時間的な対応付けを手がかりとして両話者の音声の特徴空間の対応付けを求め、これを用いて声質変換を行なうものである。つまり本方式では、ある話者が発声したいくつかの学習単語を用いて、規則合成音や他の人の音声の声質をその話者に近い声質に変換できる。特徴量としてスペクトルパラメータ、ピッチ周波数、及び音声のパワーを取り上げ、各特徴量をそれぞれ独立にベクトル量子化、またはスカラ量子化して、変換コードブックを作成している。男声-女声間の声質変換と男声-男声間の声質変換について、聞き取り実験を行なった結果、ABX法による実験では、男女声の変換では確実に変換先の声質であると判定され、男声間の変換でも65%の割合で変換先の話者の声であると判定された。また、一対比較による実験結果を数量化4類で解析したところ、変換された音声が発換先に近い声質を持つと判断されていることが確認された。

文献：[Abe,M-87-10][M.Abe-88-02][M.Abe-88-03][M.Abe-88-04][M.Abe-89-03][Abe,M-89-5][Abe,M-90-3a]

12.1.2 異言語間の声質変換

自動翻訳電話の入力は、日本人の発声した日本語であり、出力は翻訳された英語である。言語間にわたる声質変換の目的は、この出力英語に日本人の個人性を付与することである。すなわち言語間にわたる声質変換とは、音声の個人性を保ちつつ、発声した言語と異なる言語の音声を生成することである。この問題に対して2つの検討を行なった。第一に、バイリンガル話者が発声した日本語と英語を用いて、言語の違いによるスペクトルの差を調べた。その結果(1)言語間のスペクトルの差は、個人間のスペクトルの差よりも小さいこと、(2)言語間の静的なスペクトルの差は、聞こえの点であまり重要でないことが明らかとなった。第2に、言語間にわたる声質変換のモデルとモデルのための尺度を提案する。声質変換は、2話者のコードブックのマッピング問題として扱い、コード間の対応付けを相互情報量とエントロピーによって定量化する。男声の英語規則合成音に日本人女性の声質を付与する実験を行なった結果、英語の音質を保ちながら女声の声質をもつ合成音が得られた。

文献 : [Abe,M-90-2][Abe,M-90-2][Abe,M-90-3b][Abe,M-90-4][Abe,M-90-6][Abe,M-90-11][Abe91JASA7]

12.1.3 セグメントを単位とする声質変換

音声の個人性に関して、音素セグメントには、瞬時のスペクトル特徴ばかりではなく、スペクトルの連なりで決定される動的特徴が含まれている。音声に含まれる個人性制御の高度化を図るために、音素セグメントを変換の単位とする声質変換方式について検討した。変換対象となる 2 人の話者は同一内容のリストを発声し、この音声を用いて、話者間で音素セグメントの対応関係が決定される。変換時には、入力音声は音声認識モジュールによって音素セグメントに分割され、分割された各音素セグメントは、音素セグメントの対応関係に基づいてターゲットとなる話者の音素セグメントに入れ替えられる。本方式を 2 名の男性話者間で、25 文の音声を用いて変換実験を行なった。ターゲット話者の音声と変換後の音声のスペクトル歪みは、ターゲット話者の音声と変換前の音声とのスペクトル歪みに比べて、1/3 に減少した。また、聞き取り実験の結果、セグメント単位で変換された音声は、フレーム単位で変換された音声に比べて、話者識別の点で 20% の向上が得られた。

文献 : [Abe,M-90-9a][Abe91SP2][Abe91ASJ3][Abe91ICASSP5]

第 13 章

自動翻訳電話

13.1 自動翻訳電話 (音声翻訳) の研究

ATR 自動翻訳電話研究所の研究目標は、自動翻訳電話 (あるいは音声翻訳) の研究である。

従来までの通信技術は、「いつでもどこでも誰とでも情報をそのままやりとりできる」ことを大きな目標としてきた。かつては全国即時通話により「いつでもどこでも誰とでも」話せることが電話の目標であった。テレビ電話やファクシミリは、音声以外により多くの情報を、変形させずに送ることが大きな目標であった。光ファイバや衛星通信などの技術も、情報を変形せずに大容量の伝送を目指していた。一方、符号化技術は、情報をできるだけ変形せずに情報を圧縮することに努力が払われてきた。

しかし、このように距離の壁を克服しても、人々間のコミュニケーションがスムーズにできるようになったわけではない。残された最大の問題の一つが、言語の壁である。これを、技術によって解決しようとするのが自動翻訳電話技術である。

自動翻訳電話を最初に提唱したのは日本電気である。その後、ATR 自動翻訳電話研究所が設立されて、公共的な性格の大きな規模の研究が開始されてから、世界的に波及するようになった。現在、この分野の研究を行なっている研究機関はまだ十指に満たないが、今後急速に増加するものと予想されている。なお、この分野は、電話に限らないので、テキストの翻訳と対照させて、「音声翻訳」(speech translation) という語が使われつつある。

自動翻訳電話は、現在のところ、音声認識・言語翻訳・音声合成の3部分から構成される。将来は、音声波形から直接他の言語の音声波形に変換されることもあるかも知れないが、当分はこのような原理に基づくことになる。

13.1.1 自動翻訳電話実験システム SLTRANS

ATR 自動翻訳電話研究所において、最初に構築された自動翻訳電話実験システムは、“SLTRANS”(Spoken Language Translation System) と呼ばれた。これは、計算機内にあらかじめ蓄えられた音声波形データを特定話者モードで認識して、言語翻訳するものであった。

そこで用いられた音声認識部の方式は、音素モデルとしてはファジィベクトル量子化複数コードブック HMM、構文解析としては一般化 LR パーザであった。但し、オンラインシステムではなく、文節ごとにあらかじめ区切ってディスクファイルとして蓄積されていたものを入力して認識するものであった。文節ごとに複数の認識候補が出力され係り受け解析部に送られた。係り受け解析部は、統計的な情報を用いて、文節ラティスからもっともらしい文章をつくり出すものであった。翻訳部は単一化文法に基づくものであった。このシステムは、1990年11月の研究発表会にて公開された。

13.1.2 自動翻訳電話実験システム SLTRANS2

マイクロフォンからのオンライン入力で、高速な音声認識処理を可能となるよう SLTRANS を改良システムとして、SLTRANS2 を構築した。音声入力部に、試作した専用ハードウェアである HMM-LR 方式大語彙

表 13.1: 音声認識率および翻訳率 (話者適応 25 単語, LR パーザのビーム幅上限 250, 評価データ 259 文)

話者	文節間の処理	文節認識率 (%)			文認識率 (%)			英語への翻訳率 (%)
		1 位	~ 2 位	~ 3 位	1 位	~ 2 位	~ 3 位	
MIK	事後処理	91.3	95.9	97.0	85.7	90.7	91.1	90.4
	逐次予測	91.5	96.2	97.2	85.3	91.1	91.9	90.7
MST	事後処理	87.3	93.0	95.9	76.4	83.4	86.5	86.5
	逐次予測	87.3	93.2	95.6	76.8	84.6	88.4	86.1
FAK	事後処理	94.3	97.0	98.3	89.2	92.7	93.8	90.7
	逐次予測	95.3	97.4	98.6	90.7	94.2	95.0	91.0
FNY	事後処理	87.8	96.2	97.3	77.6	89.2	91.5	86.1
	逐次予測	88.0	95.5	96.4	77.6	90.3	91.9	86.1

連続音声認識装置を使用した。方式はやはりファジィベクトル量子化複数コードブック HMM である。また、この装置のために、専用のファジィコードブックマッピング方式による話者適応の装置も開発した。このシステムでは、係り受け解析部ははずし、かわりに文節間文法を適用するポストフィルタを用いた。このシステムは、1991 年 11 月の研究発表会にて公開された。

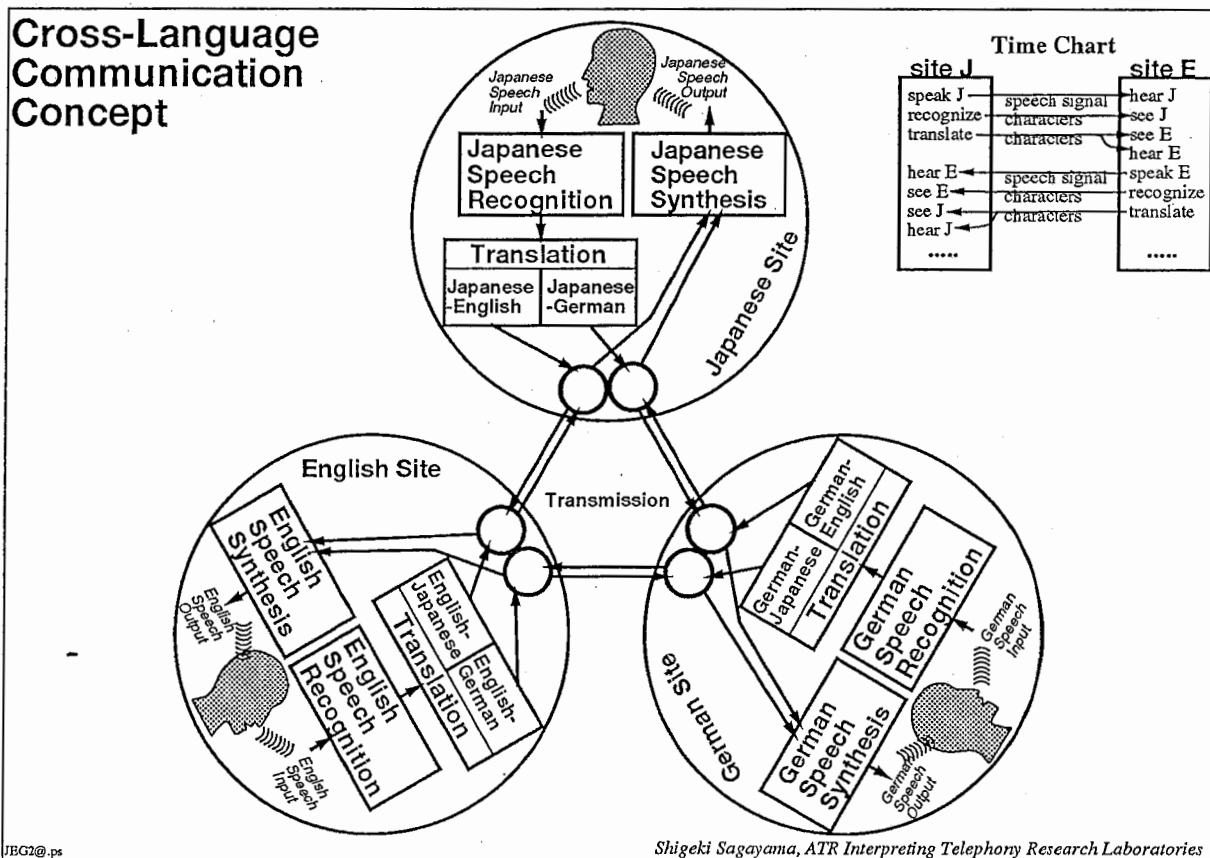


図 13.1: 自動翻訳電話接続

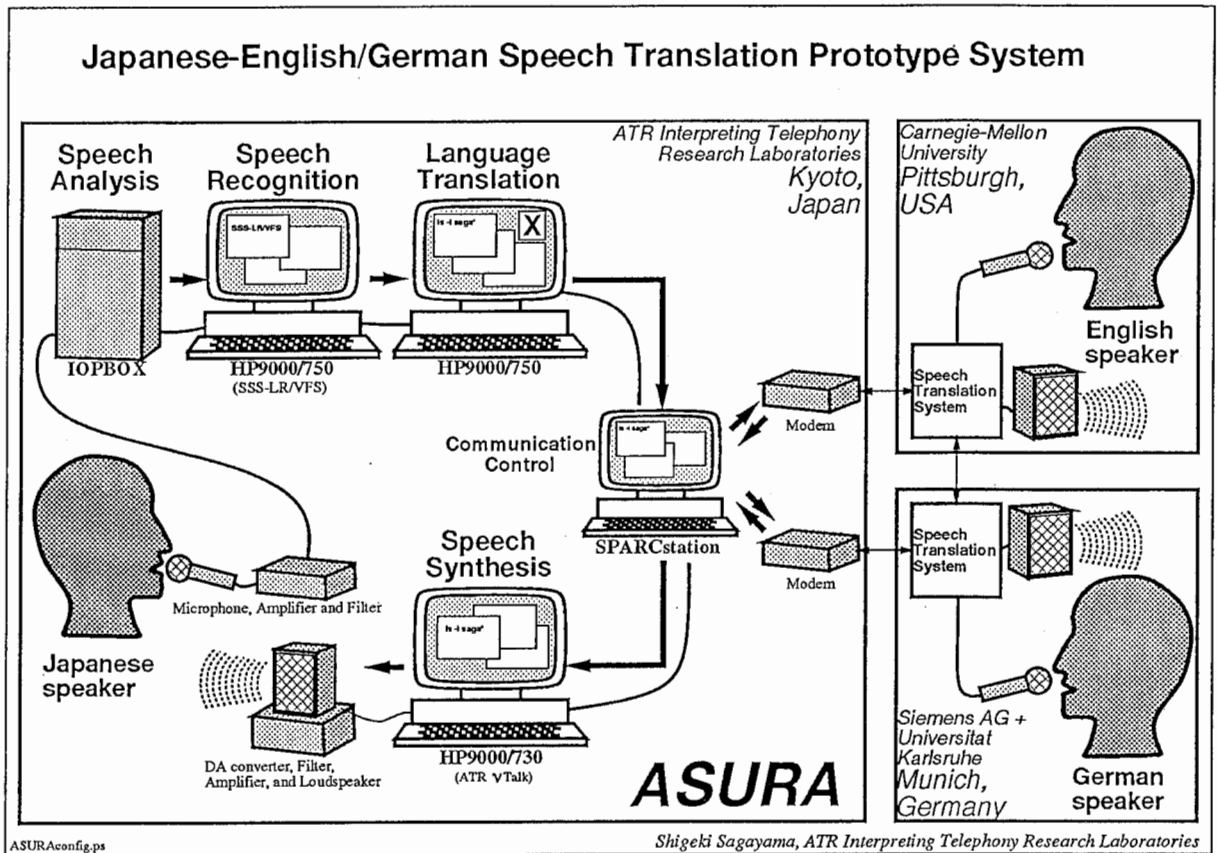


図 13.2: ASURA の構成

13.1.3 自動翻訳電話実験システム ASURA

ATR 自動翻訳電話研究所における自動翻訳電話実験の最終システムを、“ASURA”(Advanced Speech Understanding and Rendering system of ATR) と呼ぶ。このシステムは、日本語音声から英語およびドイツ語の音声に翻訳するもので、1992年11月の研究発表会にて公開された。

ASURA のために最終的に採用した連続音声認識方式の要点を以下に示す。

1. 音声分析

音声のサンプリング周波数 12kHz、16次 LPC 分析、音声パラメータは 16次 LPC ケプストラム、16次デルタ LPC ケプストラム、対数パワー、デルタ対数パワーの計 34次元

2. 音素モデル

単一ガウス分布 600 状態の隠れマルコフネットワーク (約 1,700 種の音素環境クラスタを表現する)、音素環境依存の音素継続時間制御

3. 文法

語彙約 1,000 語の文脈自由文法

4. 構文解析と解探索

文節ごとに音素環境依存型一般化 LR 構文解析法、可変幅ビーム探索、文全体としては二段階 LR 構文解析

性能を以下に示す。

13.1.4 自動翻訳電話国際実験 (CSTAR 実験)

ATR 自動翻訳電話研究所では、米国のピッツバーグのカーネギーメロン大学、ドイツのミュンヘンのシーメンス社およびカールスルーエのカールスルーエ大学と協同で、自動翻訳電話の研究に関して協力関係を持ち、CSTAR (Consortium of Speech Translation Advanced Research) を作って共同実験や共同開発を行なった。1993年1月28日には、日・米・独の3か国を結んで自動翻訳電話の公開の国際実験が初めて行なわれ、成功した。

CSTAR で採用した自動翻訳電話相互接続の構成は、 のようである。このような構成法の特徴は、各国側とも、主に次のような3つの部分からなりたっていることである。

- **音声認識** …… 音声で話された言葉を、文字の表現に変換する。
- **言語翻訳** …… 文字で表現された自国語を相手国言語に変換し、デジタル回線を通して相手国に送る。
- **音声合成** …… デジタル回線を通して相手国から受けとった自国語の文字表現を、音声に変換する。

つまり、日本側 (ATR) には、図に示すように、日本語音声認識、日英および日独言語翻訳、日本語音声合成がある。その日本側の構成は、以下のようである。このようにする理由は、各国とも、音声認識と音声合成は、自国語の分だけ持てばよく、研究もしやすいからである。

さらに、一方は音声入出力で、一方はキーボードと表示画面を用いて会話をすることも可能である。これは、音声認識と合成がまだ進歩していない国との通信も可能にする。また、言語翻訳部分も、国際ネットワークのどこにおいてもよいわけであるから、言語翻訳部分を必ずしも自国で持たなくとも翻訳通信が可能である。このように、この方式によると早い段階から音声翻訳通信を実用にもすることも可能である。

第 14 章

音声データベース

14.1 音声データベース

音韻ラベル付きの音声データベースは音声認識の研究のみならず音声合成の研究にも必要不可欠である。

14.1.1 大規模日本語音声データベース

他のいくつかの研究機関とも協力して、大規模日本語音声データベースを構築した。単語音声データベース (セット A) の作成から開始し、連続音声データベース (セット B)、不特定話者用音声データベース (セット C) の作成を行なった。各データの内容については後述する。音声データを種々の研究目的に効率良く利用するため、1) 音声記号層、2) イベント層、3) 異音化層、4) 融合化層、5) 母音中心層の 5 階層からなる階層的な音韻ラベルを用いて音韻環境を記述した ([Takeda-87-8] [Sagisaka-87-5] [Takeda-87-3] [Takeda-87-6] [Kuwabara-87-4-b] [Takeda-88-01b] [Takeda-88-05] [Kurematsu-88-05] [Takeda-88-10c] [Kuwabara-89-05] [Kuwabara-89-07c] [Kurematsu-89-09] [Kurematsu 90-3] [Abe,M 90-9b] [Sagisaka 90-11])。

連続音声データベースについては、できるだけ少ない数の文で実際の日本語に含まれる発声を可能な限りカバーするようにするため、発声現象の基本単位として音素をとり、その単位の出現頻度に関してバランスのとれた文セットを作成した ([Iso-88-03])。また、連続音声データベースに対しては、音韻ラベルに加え、基本周波数、言語情報、韻律情報も付加した ([M.Abe-89-04a] [M.Abe-89-04b] [M.Abe-89-10])。

作成した音声データベースは CD-ROM として販売され、日本における事実上の標準データベースとして利用されている。データベースの販売については後で詳しく述べる。

14.1.2 英語音声データベース

自動翻訳電話では、日英の翻訳システムを想定しているため、英語音声も研究対象に含まれる。そのためには、英語の音声データベースが必要になってくる。国内での英語音声の収集とラベル付与には多大の困難が予想されるため、エジンバラ大 (英国) およびカーネギーメロン大 (米国) との共同作業という形でデータベースの作成を行なった。

14.1.3 データベース管理システム

蓄積された音声データの検索・提供を容易にするため、階層的音韻ラベルを基にした音声データベース管理システムを構築した。システムは関係データベース UNIFY 上に構築され、データアクセスのためのインターフェースとして簡易検索言語 EAL を作成した ([Takeda-87-8] [Takeda-87-9] [Takeda-87-12] [Takeda-88-01a] [Takeda-88-10c])。

14.1.4 音声ワークベンチ

研究用の音声ワークベンチを micro Vax station 上の X11 ウィンドウ上に構築した。高品質のスペクトログラムと機能の拡張性、データへのアクセスの容易さが特徴である ([Maruyama-87-12a] [Maruyama-87-12b] [Maruyama-88-01] [Maruyama-88-02] [Maruyama-88-03] [Murayama-88-10] [Cheveigné-89-10b])。

14.1.5 各音声データのあらまし

Set A (大語彙単語音声および文節・文章音声データベース)

セット構成 :	5240 単語『新明解国語辞典より抜粋』重要語	(5240 個)
	音素連鎖バランス単語リスト	(216 個)
	101 音節『単音節』	(101 個)
	数字 A・B	(15 + 10 = 25 個)
	アルファベット	(35 個)
	外来音	(9 個)
	文章『国際会議の申し込み』3 発声様式	(115 文 × 3 発声)

発声方法 : 文章データについては、
文節A発声『長単位』(SA)
文節B発声『短単位』(SB)
文発声『区切りなし』(SC)

話者 : アナウンサー 16名 (男性8名 女性8名)
ナレータ 4名 (男性2名 女性2名)

話者名 : アナウンサー (男声: MAU, MNM, MMS, MMY, MXM, MTT, MTM, MSH;
女声: FSU, FKS, FFS, FMS, FYN, FKM, FAF, FTK)
ナレータ (男声: MHT, MTK; 女声: FKN, FYM)

発声回数 : 2回

ラベリング : 1回目発声のみ

音声切り出し : 全セット、1、2回目とも

Down sampling : 1回目発声のみ

Set B (音素バランス文音声データベース)

セット構成 : 503文『A~J』(503文 x 2)

発声方法 : 連続発声『区切りなし』
文節発声『区切りあり』

話者 : ナレータ 8名 (男性4名 女性4名)
アナウンサー 3名 (男性2名 女性1名)

話者名 : ナレータ (MHT MYI MHO MTK FKN FYM FKS (FKN))
アナウンサー (MMY MSH FTK)

発声回数 : 2回*

ラベリング : MYI・MHO・MTK・FYM・FKS・FKN** : 自由発声のみ (6名)
MHT・FKN・MMY・MSH・FTK 両発声とも

音声切り出し : 1回目発声のみ

Down sampling : 連続発声のみ (1回目発声)

* ... MYI・MHO・MTK・FYM・FKS・FKNは1回のみ発声となっている。

** ... 両発声ともラベリングされているFKNと同一話者であるため、外部販売はしない方針である。

Set C

Cセット 520 単語『5240 単語からの抜粋』(520 個)

音素連鎖バランス単語リスト (216 個)

数字A (15 個)

文章『Bセットからの抜粋・A～C』 (150 文 x 2)

発声方法 : 文章データ

連続発声『区切りなし』

文節発声『区切りあり』

話者 : 素人

発声回数 : 1 回

ラベリング : 文章データは、連続発声のみ付与

(ATRでは、両発声とも付与。他社の一部に両発声付与あり)

音声切り出し : 全社、全データ

Down_sampling : 全データ

※ 当データは、現在収集中 (平成4年まで)

Set D

セット D は、主として音声合成研究のために収集されたデータである。

《単語》	CV b、CV g、CV k、CV s、CV m、単語 (A_ 6) 発声方法 : 8 通りの発声 (男性) (normal,speed_fast,speed_slow,speed_norm,pitch_low, pitch_high,power_low,power_high) 6 通りの発声 (女性) (normal,speed_fast,speed_slow,speed_norm,power_low, power_high) CV s、CV m、単語 (A_ 6) については、男女とも 6 通りの発声	
《単音》	101 音節 キャリア付き単音節 (これは)	(101 単音節) (109 音語)
《文章》	文セット_ 1 文セット_ 2(for_ALEX) 文セット_ 3(2 文節) 文セット_ 4(3 文節) 3 文節抜粋 やさしいバイオテクノロジー 機械との共存 すずめと人間 いよめし 北国の園芸だより 沖縄のもずく料理 芹そば 春が来ると 九州 関東 ヨーロッパ アメリカ	(215 文) (3 文) (825 文) (346 文) (141 文) → 5 通りの発声 (speed fast, speed slow, normal x 3) (33 文) (22 文) (25 文) (12 文) (37 文) (17 文) (12 文) (115 文) (29 文) (28 文) (37 文) (43 文)
《会話文》	1 (11 文) 7 通りの発声 (normal,speed fast,speed slow,speed norm,okori,isogi,shinsetsu) 2 (7 文) 8 通りの発声 (normal,speed fast,speed slow,speed norm,okori,isogi,shinsetsu, reading) 3 (12 文) 8 通りの発声 (normal,speed fast,speed slow,speed norm,okori,isogi,shinsetsu, reading)	

話者 : ナレータ 2 名 (男性 1 名 女性 1 名)

話者名 : ナレータ (MHT FKN)

発声回数 : 1 回

ラベリング : 全データ

(会話文、2 文節、3 文節、文セット_ 1 のみモーラ中心ラベルあり)

音声切り出し : 全データ

Down_sampling : 全データ

Set F

Fセットは、1,122文の認識最終性能評価用データである。

構成	音韻バランス (503 文) 外来音節等を含む (19 文) 機能試験文 (600 文)
発声方法	文節発声
話者	ナレータ 6 名 (男性 3 名 女性 3 名)
話者名	ナレータ (M01 M02 M03 F01 F02 F03)
発声回数	1 回
ラベリング	始末端
タスクコード	SH (ファイル名: xxx_ SH 0001. AD)

モデル会話文データ

1. データのあらまし

1-1. MAU・MNMについて

- ・このデータは、12セット(1～10、A、B)全てを発声。
- ・発声は3種類あり、ファイルは種類別に命名する。

通常発声… 1…MAU_M01_01. LB (NEW)
長単位発声… 2… M02
短単位発声… 3… M03

通常発声 → AセットDBのDSCに相当
長単位発声 → DSAに相当
短単位発声 → DSBに相当

1-2. MTKについて

- ・このデータは、5セットのみ発声(1～5)。
- ・発声は2種類あり、ファイルは種類別に命名する。

dialogue…D…MTK_M1_D01. LB (12K)
reading…R… R01

dialogue → パートナーを設け、事務局と質問者の役割を交替して発声
(機械処理の時点で、編集をおこなっている)

reading → パートナーは設けず、発声者が両方の役割を発声

1-3. FKN・FMPについて

- ・このデータは、12セット(1～10、A、B)全てを発声。
- ・発声の種類、命名、発声方法は2-2に同じ。

2. 評価用モデル会話データとの相違点

データ処理研究室にて収集した「モデル会話」と、当dir内の「モデル会話」には、発声リストに相違点がある。当データでは、『もしもし』『はい』『いいえ』で始まる文章は、全て句点(.)で区切られている。(連続発声/文節発声とも)しかし、評価用データの場合には、連続発声に限り句点(.)が読点(,)に変更されている。(以下に実例をあげる)

会話 No.	・連続発声	・文節発声	
1	質問者 1	もしもし、そちらは～	もしもし。そちらは～
	事務局 1	はい、そうです。	はい。そうです。
	質問者 3	いいえ、まだです。	いいえ。まだです。
2	事務局 1	はい、こちらは～	はい。こちらは～
3	事務局 1	はい、こちらは～	はい。こちらは～
4	質問者 2	いいえ、持っていません。	いいえ。持っていません。
	質問者 5	はい、そうです。	はい。そうです。
5	事務局 1	はい、こちらは～	はい。こちらは～
	質問者 3	はい、そうです。	はい。そうです。
6	事務局 1	はい、こちらは～	はい。こちらは～
7	事務局 1	はい、こちらは～	はい。こちらは～
	事務局 2	はい、何でしょうか。	はい。何でしょうか。
	質問者 5	いいえ、ありません。	いいえ。ありません。
8	事務局 1	はい、会議事務局です。	はい。会議事務局です。
	質問者 4	はい、そうです。	はい。そうです。
	事務局 5	はい、分かりました。	はい。分かりました。
9	事務局 1	はい、会議事務局です。	はい。会議事務局です。
	事務局 5	いいえ、どういたしまして。	いいえ。どういたしまして。
10	事務局 1	はい、会議事務局でございます。	はい。会議事務局で～
	質問者 6	はい、そうです。	はい。そうです。
A	質問者 1	もしもし、そちらは～	もしもし。そちらは～
	事務局 1	はい、そうです。	はい。そうです。
	質問者 3	いいえ、まだです。	いいえ。まだです。

14.1.6 データベース販売のご案内

ATRでは、音声認識、音声合成および音声知覚等の研究に用いる、大語彙の高品質な研究用日本語音声データベースを構築しており、単語音声データ(8,500語/話者)20話者、文章音声データ(10,000語/話者)2話者について、全国の多くの大学、企業の研究所等に販売を行って、ご利用をさせて頂いております。更に、単語・文章音声データ15～20話者を1セットに格納した多数話者音声データベース、自然言語処理等の研究を目的として、模擬対話実験を基に構築した対話データベースを整備しており、この度ご利用頂くべく販売を開始しました。以下に、音声データベースおよび対話データベースの内容についてご案内いたします。

構成

セット A (大語彙単語音声および文節・文章音声)

NO	話者	内容	種類	標準規格	備考
1	MAU	男性アナ	単語音声データ (8,500 語)	unix システム tar コマンド格納	話者毎に販 売します
2	MHT	男性ナレータ			
3	FKN	女性ナレータ			
4	FSU	女性アナ			
5	FKS	女性アナ			
6	FYN	女性アナ			
7	MTK	男性ナレータ			
8	MMY	男性アナ			
9	MMS	男性アナ			
10	MNM	男性アナ			
11	MXM	男性アナ			
12	FFS	女性アナ			
13	FYM	女性ナレータ			
14	FMS	女性アナ			
15	FKM	女性アナ			
16	FAF	女性アナ			
17	FTK	女性アナ			
18	MTT	男性アナ			
19	MTM	男性アナ			
20	MSH	男性アナ			

セット B (音素バランス文章音声)

51	MYI (男性アナ)	連続: ラベル付 基本周波数 ・言語韻律情報付 文節:	文章音声データ 503 文 (10,000 語)		
52	MTK (男性ナレータ)	連続: ラベル付 文節:	12kHz 連続発声のみ または 20kHz 連続発声 および文節区切発声	"	"
53	MHO (男性ナレータ)	連続: ラベル付 文節:			
54	MHT (男性ナレータ)	連続: ラベル付 基本周波数 文節: ラベル付			
55	MMY (男性アナ)	連続: ラベル付 文節: ラベル付			
56	MSH (男性アナ)	連続: ラベル付 文節: ラベル付			
57	FKN (女性ナレータ)	連続: ラベル付 基本周波数 文節: ラベル付			
58	FKS (女性ナレータ)	連続: ラベル付 文節:			
59	FTK (女性アナ)	連続: ラベル付 文節: ラベル付			
60	FYM (女性ナレータ)	連続: ラベル付 基本周波数 文節:			

セット C (多数話者)

セット C は 6 種類のデータセット (C1 ~ C6) からなり、内容については以下のとおりです。

NO	話者	内容	標準規格	備考
C1-M01	男性 20 話者	多数話者データ	unix システムの tar コマンドで格納	話者グループ毎 に販売します
C1-F01	女性 20 話者	最重要語 (520)		
C2-M01	男性 20 話者	多数話者データ	1 ブロック 512 バイト 20 ブロックで ブロック化	サンプリング 周波数 20kHz
C2-F01	女性 20 話者	バランスリスト (216) + 数字 A(15) + 連続発声 A(50)		
C3-M01	男性 20 話者	多数話者データ	MT 記録密度	
C3-F01	女性 20 話者	連続発声 B(50) + 連続発声 C(50)		
C4-M01	男性 20 話者	多数話者データ	6250bpi	
C4-F01	女性 20 話者	文節発声 A(50)	オープンリール	
C5-M01	男性 20 話者	多数話者データ	または	
C5-F01	女性 20 話者	文節発声 B(50)	12,500ftpi	
C6-M01	男性 20 話者	多数話者データ	150M データカート	
C6-F01	女性 20 話者	文節発声 C(50)	リッジ	

注 1. A ~ C は、音素バランスを考慮して選んだ 50 文章からなるセットで内容が異なる。

対話テキストデータベース

1. 対話の内容

- ・国際会議の申込に関する参加者と事務局の対話。
- ・旅行に関する旅行会社と客との対話。

2. 対話のメディア

- ・対話が行われるメディアは電話またはキーボード。

3. 使用言語

- ・対話は日本語話者同士、または、日本語話者と英語話者との間で行っています。後者の対話は、通訳を介して行っています。

4. 事前分析

- ・対話データには以下に示す事前分析を加えています。

- ◎英語対訳の付与
- ◎各種言語単位 (単語、文節、文など) への分割
- ◎単語の各種属性の付与
- ◎単語間の係り受け関係の付与
- ◎日英対応関係の付与

5. データベースの種類

- ・データは会話単位 (対話の開始から終了までが 1 単位) に収容されており、セット 1 ~ 4 に分けて提供いたします。

set	タスク	メディア	標準規格
1	国際会議	電話	unix システム
2	国際会議	キーボード (準備中)	tar コマンドで格納
3	旅行	電話	
4	旅行	キーボード	12,500ftpi 150M データカートリッジ

[参考文献] 江原暉将ほか: “ATR対話データベースの内容,” ATRテクニカルレポートTR-I-0186, 1990年10月

入手方法

連絡先

音声データベース、対話テキストデータベースの販売に関するお問い合わせ等は下記までお願いいたします。

〒619-02 京都府相楽郡精華町光台2-2
 (株) 国際電気通信基礎技術研究所
 企画部 (担当: 大坪)
 Tel (07749) 5 - 1177, Fax (07749) 5 - 1108

販売価格

以下のデータの販売価格は、単位あたり、以下の通りです。(予告なく変更されることがあります。)

表 14.1: データの販売価格

	企業	大学
単語音声データベース	60 万円	3.5 万円
文章音声データベース	35 万円	”
多数話者音声データベース	60 万円	”
対話テキストデータベース	5 万円	5 万円

大学関係には安価ですが、この場合にはメディアおよびコピー作業の料金が含まれていませんので、原則として ATR にお越し頂き、ご自分で作業を行なって頂くことになっております。

第 15 章

音声情報処理研究室に在籍した研究員のリスト

15.1 研究員リスト

ここに示す研究員のリストは、音声情報処理研究室に在籍した研究者のリストで、現在把握している情報である。記載洩れは極力防いだが、皆無ではないかもしれない。ただし、短期間の滞在や短期のアルバイトについては省略している。なお、このレポート全体は、音声情報処理研究室に限らず他の研究室(データ処理研究室と言語処理研究室)のメンバーによる、音声情報処理関連の研究活動も若干含まれているので、ここに示す組織としての音声情報処理研究室のメンバーリストとは一致しない。

氏名 Japanese name	name & title	position	期間 period
鹿野 清宏	Kiyohiro Shikano, Dr	室長	6/1986 - 1/1990
嵯峨山 茂樹	Shigeki Sagayama	室長	2/1990 - 3/1993
匂坂 芳典	Yoshinori Sagisaka, Dr	主幹研究員	4/1986 - 3/1993
桑原 尚夫	Hisao Kuwabara, Dr	主幹研究員	10/1986 - 7/1989
杉山 雅英	Masahide Sugiyama, Dr	主幹研究員	2/1990 - 3/1993
梅田 哲夫	Tetsuo Umeda	主幹研究員	7/1989 - 6/1990
川端 豪	Takeshi Kawabata, Dr	主任研究員	9/1986 - 2/1990
田村 震一	Shin-ichi Tamura	主任研究員	9/1986 - 2/1990
沢井 秀文	Hidefumi Sawai, Dr	主任研究員	4/1988 - 3/1991
武田 一哉	Kazuya Takeda	研究員	8/1986 - 2/1990
中村 哲	Satoshi Nakamura	研究員	9/1986 - 8/1989
宮武 正典	Masanori Miyatake	研究員	9/1986 - 3/1989
畑崎 香一郎	Kaichiro Hatazaki	研究員	12/1986 - 3/1989
花沢 利行	Toshiyuki Hanazawa	研究員	3/1987 - 2/1990
丸山 克輝	Katsuteru Maruyama	研究員	3/1987 - 2/1990
阿部 匡伸	Masanobu Abe	研究員	4/1987 - 2/1991
安部 勝雄	Katsuo Abe	研究員	3/1987 - 2/1990
中村 雅己	Masami Nakamura	研究員	9/1987 - 8/1990
小森 康弘	Yasuhiro Komori	研究員	9/1988 - 2/1992
服部 浩明	Hiroaki Hattori	研究員	5/1989 - 4/1992
大倉 計美	Kazumi Ohkura	研究員	9/1989 - 3/1993
海木 延佳	Nobuyoshi Kaiki	研究員	11/1989 - 10/1992
鷹見 淳一	Jun-ichi Takami	研究員	11/1989 - 3/1993
片岸 一起	Kazuki Katagishi, Dr	研究員	2/1990 - 3/1993
永井 明人	Akito Nagai	研究員	3/1990 - 3/1993
福沢 圭二	Keiji Fukuzawa	研究員	4/1990 - 3/1993
三村 克彦	Katsuhiko Mimura	研究員	5/1990 - 3/1993

藤原 紳吾	Shingo Fujiwara	研究員	5/1990 - 3/1993
岩橋 直人	Naoto Iwahashi	研究員	10/1990 - 3/1993
村上 仁一	Jin-ichi Murakami	研究員	3/1991 - 3/1993
加藤 喜永	Yoshinaga Kato	研究員	4/1991 - 3/1993
山口 耕市	Kouichi Yamaguchi	研究員	5/1991 - 3/1993
宮沢 康永	Yasunaga Miyazawa	研究員	7/1991 - 3/1993
小坂 哲夫	Tetsuo Kosaka	研究員	9/1991 - 3/1993
磯谷 亮輔	Ryosuke Isotani	研究員	4/1992 - 3/1993
A・ワイベル	Alex Waibel, Dr	客員研究員	5/1987 - 8/1988, 6 - 9/1989
斉藤 博昭	Hiroaki Saito, Dr	客員研究員	2/1988 - 7/1989
W・ポージー	William Poser, Dr	客員研究員	9/1988 - 2/1989
A・ド・シュヴェニエ	Alain de Cheveigné, Dr	客員研究員	2/1989 - 2/1990
D・フーバー	Dieter Huber, Dr	客員研究員	3 - 6/1990
D・レイントン	David Rainton, Dr	客員研究員	4/1990 - 3/1993
N・キャンベル	Wilhelm N. Campbell, Dr	客員研究員	8/1990 - 3/1993
H・シンガー	Harald Singer	客員研究員	3/1991 - 3/1993
H・ルッケ	Helmut Lucke, Dr	客員研究員	4/1992 - 3/1993
王文俊	Wang Wern-jun	客員研究員	2 - 7/1991
J・ヴェシエール	Jacqueline Vaissiere, Dr	客員研究員	7 - 8/1992
P・バグショー	Paul Christopher Bagshaw	客員研究員	5 - 11/1992
村山 浩一	Kouichi Murayama	研究技術員	9/1987 - 8/1988
田中 孝明	Takaharu Tanaka	研究技術員	9/1988 - 8/1989
北垣内 浩二	Kouji Kitagaito	研究技術員	9/1989 - 8/1990
荒木 忍	Shinobu Araki	研究技術員	9/1990 - 8/1991
小野 佳範	Yoshinori Ono	研究技術員	9/1991 - 3/1993
H・セゴ	Hubert Segot	学外実習生 (ENST)	4 - 9/1987
北川 英一郎	Eiichiro Kitagawa	学外実習生 (早稲田大学)	8/1987
橋本 賢治	Kenji Hashimoto	学外実習生 (早稲田大学)	8/1987
山崎 岳人	Takahito Yamazaki	学外実習生 (静岡大学)	8/1988
F・ムラド	Furati Mulad	学外実習生 (静岡大学)	8/1988
P・ハフナー	Patrick G. Haffner	学外実習生 (ENST)	3 - 11/1988
橋本 泰秀	Yasuhide Hashimoto	学外実習生 (豊橋技術科学大学)	7 - 8/1987
遠藤 隆	Takashi Endo	学外実習生 (早稲田大学)	8/1987
古川 令	Rei Furukawa	学外実習生 (早稲田大学)	3 - 4/1989
J-C・ダン	Jean-Claude Dang	学外実習生 (ENST)	3 - 11/1989
平田 好充	Yoshimitsu Hirata	学外実習生 (豊橋技術科学大学)	1 - 2/1989
野田 充	Mitsuru Noda	学外実習生 (豊橋技術科学大学)	1 - 2/1989
南 泰浩	Yasuhiro Minami	学外実習生 (慶応大学)	3-6/1989, 8/1989-2/1990, 7-10/1990
伊藤 克旦	Katsunobu Ito	学外実習生 (東京工業大学)	10 - 12/1989
上田 佳央	Yoshio Ueda	学外実習生 (豊橋技術科学大学)	1 - 2/1990
平戸 毅	Takeshi Hirado	学外実習生 (豊橋技術科学大学)	1 - 2/1990
中村 悟	Satoru Nakamura	学外実習生 (慶応大学)	7 - 9/1990, 3 - 4/1991
野村 俊之	Toshiyuki Nomura	学外実習生 (名古屋大学)	7 - 8/1990
渡戸 守	Mamoru Watado	学外実習生 (早稲田大学)	7 - 8/1990
稲垣 英浩	Hidehiro Inagaki	学外実習生 (早稲田大学)	7 - 8/1990
丸山 康	Yasushi Maruyama	学外実習生 (信州大学)	3-4/1990
A・ビエム	Alain Biem	学外実習生 (INT)	8/1990 - 1/1991

甲斐 充彦	Atsuhiko Kai	学外実習生 (豊橋技術科学大学)	1 - 2/1991
菊池 英明	Hidefumi Kikuchi	学外実習生 (早稲田大学)	8/1991
田本 真詞	Masafumi Tamoto	学外実習生 (東京工業大学)	8/1991
土井 啓輔	Keisuke Doi	学外実習生 (竜谷大学)	8/1991 - 9/1992
吉田 哲也	Tetsuya Yoshida	学外実習生 (慶応大学)	8 - 9/1991
栗並 賢太郎	Kentaro Kurinami	学外実習生 (慶応大学)	8 - 9/1991
R・ブリュニャス	Romain Brunias	学外実習生 (INT)	8/1991 - 1/1992
R・ランガーニュ	Richard Lengagne	学外実習生 (INT)	8/1991 - 1/1992
岡本 匡史	Tadashi Okamoto	学外実習生 (豊橋技術科学大学)	1 - 2/1992
安芸 聡	Satoshi Aki	学外実習生 (豊橋技術科学大学)	1 - 2/1992
渡辺 秀行	Hideyuki Watanabe	学外実習生 (北海道大学)	2 - 4/1992
山本 寛樹	Hiroki Yamamoto	学外実習生 (早稲田大学)	3 - 9/1992
I・ドネスク	Ioana Donescu	学外実習生 (INT)	8/1992 - 1/1993
E・ウィレムズ	Edward David Peter Willems	学外実習生 (ENST)	8 - 12/1992
水野 理	Masashi Mizuno	学外実習生 (早稲田大学)	7 - 9/1992
F・マルタン	Franck Martin	学外実習生 (Univ de Centre Paris)	8 - 9/1992
岩井 康雄	Yasuo Iwai	Part-time (京都大学)	8/1989 - 3/1990
定延 利之	Toshiyuki Sadanobu	Part-time (京都大学)	8/1989 - 3/1990
吉田 夏也	Natsuya Yoshida	Part-time (京都大学)	8/1989 - 3/1990
栗原 美和子	Miwako Kurihara	Part-time (同士社大学)	4/1990 -
M・トロニエ	Mechtild Tronnier	Part-time (関西大学)	9/1990 - 4/1991
周 明勇	Ming-Yong Zhou	Part-time (大阪市立大学)	10/1990 -
衣笠 和江	Kazue Kinugasa	Part-time (関西大学)	2/1992 - 3/1993

第 16 章

文中で示した文献のリスト

16.1 文献リスト

以下に掲載する文献リストは、毎年まとめてきた外部発表論文集“Research Activities at Speech Processing Department”の目次の部分を集めたものである。文中で示した文献名はこの目次に記載されている略記に従っており、その略記文献名から論文のタイトル、著者、学会名、掲載ページなどの情報を検索することができる。

Technical Publication List in Speech Processing Department
(April, 1986 through October, 1987)

(Bold letter titles show papers written in English or with an English abstract.)

Ref. ID	Title	Authors	Conference / Journal	Page
Shikano-86-9	カーネギー・メロン大学における音声認識・理解研究の現状 (Present State of Research Activities of Automatic Speech Recognition and Understanding at Carnegie-Mellon University)	Seiichi Nakagawa (Toyohashi Univ. of technology), Kiyohiro Shikano	Journal of Acoustic Society of Japan, (1986-09)	15 ~ 19
Shikano-86-12-a	音声理解研究の動向 (Trends on Speech Understanding Researches)	Kiyohiro Shikano, Akira Kurematsu	Journal of Acoustic Society of Japan, (1986-12)	20 ~ 24
Shikano-87-4	Improvement of Word Recognition Results by Trigram Models	Kiyohiro Shikano	Proceedings on ICASSP 86, 29.2, (1987-04)	25 ~ 28
Sagisaka-86-9	日本語テキストからの音声合成とアクセント (Speech Synthesis by Rule and Prosodies)	Yoshinori Sagisaka, Hirokazu Sato (NTT Electrical Communication Laboratories)	Phonetics Meeting in Kansai Area, (1986-09)	29 ~ 50
Shikano-86-12-b	ベクトル量子化による話者適応 (Speaker Adaptation through Vector Quantization)	Kiyohiro Shikano	IEICE Technical Report, SP86-65, (1986-12)	51 ~ 61
Kuwabara-86-11	音声の声質変換における信号処理 (Speech Processing in Speech Quality Change)	Hisao Kuwabara, Tohru Takagi (NIHK Science and Technical Laboratories)	The First Digital Signal Processing Symposium, B2.3, (1986-11)	62 ~ 68

76

Technical Publication List in Speech Processing Department
(April, 1986 through October, 1987)

(Bold letter titles show papers written in English or with an English abstract.)

Ref. ID	Title	Authors	Conference / Journal	Page
Kuwabara-86-12	分析合成による声質変換と喋声音改善への応用 (Speech Quality Control by the Analysis-Synthesis Method and Application to the Enhancement of Abnormal Speech)	Hisao Kuwabara, Tohru Takagi (NIHK Science and Technical Laboratories)	IEICE Technical Report, SP86-57/H-86-56, (1986-12)	69 ~ 77
Kawabata-87-4	Word Spotting Method Based on Top-Down Phoneme Verification	Takeshi Kawabata, Masaki Kohda (NTT Electrical Communication Laboratories)	Proceedings on ICASSP 86, 34.7, (1987-04)	78 ~ 81
Takeda-87-8	Acoustic-Phonetic Labels in a Japanese Speech Database	Kazuya Takeda, Yoshinori Sagisaka, Shigeru Katagiri	European Conference on Speech technology, pp13-16, (1987-08)	82 ~ 85
Sagisaka-87-5	Phonetic Labeling and Acoustic Correlates for Building Japanese Speech Data Base	Yoshinori Sagisaka, Shigeru Katagiri, Kazuya Takeda	ASA-Meeting '87, Spring, (1987-05) including its handout.	86 ~ 91
Shikano-87-3	Improvement of Word Recognition Results by Trigram Models	Kiyohiro Shikano	ASJ-Meeting '87 Spring, 3-5-1 (1987-03)	92 ~ 93
Takeda-87-3	音声データベース構築のための音韻ラベリング (Acoustic-Phonetic Labeling in a Japanese Speech Database)	Kazuya Takeda, Yoshinori Sagisaka	ASJ-Meeting '87 Spring, 2-5-10 (1987-03)	94 ~ 95

Technical Publication List in Speech Processing Department
(April, 1986 through October, 1987)

(Bold letter titles show papers written in English or with an English abstract.)

Ref. ID	Title	Authors	Conference / Journal	Page
Kuwabara-87-3	声の明瞭性を支配する物理的特徴量の抽出 (Extraction of Acoustic Features Specifying the Clearness of Natural Speech)	Tohru Takagi (NHK Science and Technical Laboratories), Hisao Kuwabara	ASJ-Meeting '87 Spring, 2-6-15 (1987-03)	96 ~ 97
Kuwabara-87-4-a	声の個人性に関する諸問題 (Some Problems on the Personal Characteristics of Speech)	Hisao Kuwabara	Journal of IECEJ, Vol. 70, No. 4, (1987-04)	98 ~ 105
Kuwabara-87-8	Quality Control of Speech by Modifying Formant Frequencies and Bandwidths	Hisao Kuwabara, Tohru Takagi (NHK Science and Technical Laboratories)	the 11th International Congress of Phonetic Sciences, Tallin Estonia, (1987-08)	106 ~ 109
Kurematsu-87-4	自動翻訳電話のための研究課題—音声を主体として— (Research for Realization of Interpreting Telephone System)	Akira Kurematsu	IEICE Technical Report, Invited Talk, (1987-04)	110 ~ 117
Kuwabara-87-4-b	日本語音声データベースとそのラベリング (Japanese Speech Database and Its Phoneme Labels)	Hisao Kuwabara, Yoshinori Sagisaka, Kazuya Takeda, Shigeru Katagiri	Phonetics Meeting in Kansai Area, (1987-04)	118 ~ 123
Shikano-87-6	Improvement of Word Recognition Results by Trigram Models	Kiyohiro Shikano	IEICE Technical Report, SP87-23, (1987-06)	124 ~ 132

Technical Publication List in Speech Processing Department
(April, 1986 through October, 1987)

(Bold letter titles show papers written in English or with an English abstract.)

Ref. ID	Title	Authors	Conference / Journal	Page
Takeda-87-6	音韻ラベルを持つ日本語音声データベースの構築 (Construction of an Acoustically-Phonetically Transcribed Japanese Speech Database)	Kazuya Takeda, Hisao Kuwabara, Yoshinori Sagisaka, Shigeru Katagiri	IEICE Technical Report, SP87-19, (1987-06)	133 ~ 141
Nakamura-87-6	ベクトル量子化を用いたスペクトログラムの正規化 (Spectrogram Normalization Based on Vector Quantization)	Satoshi Nakamura, Kiyohiro Shikano	IEICE Technical Report, SP87-17, (1987-06)	142 ~ 150
Hatazaki-87-9	連続音声中の音韻認識エキスパートシステムの検討 (An Expert System for Phoneme Recognition in Continuous Speech)	Kaichiro Hatazaki, Shin'ichi Tamura, Takeshi Kawabata, Kiyohiro Shikano	IPS Fall Meeting, 1L-4, (1987-09)	151 ~ 152
Takeda-87-9	音声データベース管理システムの構築 (Construction of a Japanese Speech Database Management System)	Kazuya Takeda, Hisao Kuwabara, Shogo Morikawa (TIS)	IPS Fall Meeting, 2H-4, (1987-09)	153 ~ 154
Miyatake-87-9	会話文音声合成のための音声合成 (Speech Synthesis by Rule for Conversation)	Masanori Miyatake, Yoshinori Sagisaka	IPS Fall Meeting, 2H-3, (1987-09)	155 ~ 156
Takeda-87-10	母音無声化の要因分析と予測手法の検討 (Analysis and Prediction of Devocalized Phenomena)	Kazuya Takeda, Hisao Kuwabara	ASJ Fall Meeting, 3-3-8, (1987-10)	157 ~ 158

Technical Publication List in Speech Processing Department
(April, 1986 through October, 1987)

(Bold letter titles show papers written in English or with an English abstract.)

Ref. ID	Title	Authors	Conference / Journal	Page
Nakamura-87-10	ベクトル量子化を用いたスペクトログラムの正規化 (Spectrogram Normalization Using Vector Quantization)	Satoshi Nakamura, Kiyohiro Shikano	ASJ Fall Meeting, 3-3-8, (1987-10)	159 ~ 160
Sagisaka-87-10	種々の複合音声単位からの音声合成 (Speech Synthesis by Rule Using Non-Uniform Phonemic Clusters)	Yoshinori Sagisaka	ASJ Fall Meeting, 3-3-8, (1987-10)	161 ~ 162
Kawabata-87-10	日本語スペクトログラム特徴の英語との比較 (Sound Spectrogram Features in Japanese)	Takeshi Kawabata, Shin'ichi Tamura, Kaichiro Hatazaki	ASJ Fall Meeting, 3-3-8, (1987-10)	163 ~ 164
Miyatake-87-10-a	発声様式の違いが韻律パラメータに与える影響の分析 (Acoustic Manifestation of Prosody Control)	Masanori Miyatake, Yoshinori Sagisaka	ASJ Fall Meeting, 3-3-8, (1987-10)	165 ~ 166
Abe,M-87-10	ベクトル量子化による音質変換 (Voice Conversion through Vector Quantization)	Masanobu Abe, Satoshi Nakamura, Kiyohiro Shikano, Hisao Kuwabara	ASJ Fall Meeting, 3-3-8, (1987-10)	167 ~ 168
Hanazawa-87-10	Hidden Markovモデルを用いた日本語有声破裂音の識別 (Discrimination of Japanese Voiced Stops Using Hidden Markov Models)	Toshiyuki hanazawa, Takeshi Kawabata, Kiyohiro Shikano	ASJ Fall Meeting, 3-3-8, (1987-10)	169 ~ 170
Abe,K-87-10	抽出環境が音節スペクトルに与える影響の分析 (Spectrum Analysis by Syllable Environments)	Katsuo Abe, Yoshinori Sagisaka	ASJ Fall Meeting, 3-3-8, (1987-10)	171 ~ 172

Technical Publication List in Speech Processing Department
(April, 1986 through October, 1987)

(Bold letter titles show papers written in English or with an English abstract.)

Ref. ID	Title	Authors	Conference / Journal	Page
Miyatake-87-10-b	種々の発声様式における韻律パラメータの性質について (On the Prosodic Characteristics in Various Utterances with Different Speaking Styles)	Masanori Miyatake, Yoshinori Sagisaka	IEICE Technical Report, SP87-62, (1987-10)	173 ~ 181

Technical Publication List at the Speech Processing Department
(November, 1987 through December, 1988)

(Bold letter titles show papers written in English or with an English abstract)

Ref.ID	Title(題名)	Authors (著者)	Conference/Journal (発表先)	Page
Haffner -88-10	Fast Back-Propagation Learning Methods for Neural Networks in Speech	P. Haffner A. Waibel K. Shikano	ASJ Fall Meeting, Hakata, (1988-10)	1
Haffner -88-11a	DyNet, a Fast Program for Learning in Neural Networks	P. Haffner	ATR Technical Report, TR-1-0059, (1988-11)	—
Haffner -88-11b	Fast Back-Propagation Learning Methods for Neural Networks in Speech	P. Haffner A. Waibel H. Sawai K. Shikano	ATR Technical Report, TR-1-0058, (1988-11)	—
Hanazawa -87-12	Hidden Markovモデルを用いた日本語有声破裂音の識別 (Recognition of Japanese Voiced Stops Using Hidden Markov Models)	T. Hanazawa T. Kawabata K. Shikano	IEICE Technical Report SP87-98 (1987-12)	3
Hanazawa -88-01	Hidden Markov Modelを用いた日本語有声破裂音の識別 (Phoneme Recognition of Japanese Voiced Bursts by HMM Phone Modeling)	T. Hanazawa T. Kawabata K. Shikano	ATR Technical Report TR-1-0018 (1988-01)	—
Hanazawa -88-03	HMMを用いた音韻認識における出力確率の平滑化手法の検討 (Output Probability for HMM Phoneme Recognition)	T. Hanazawa T. Kawabata K. Shikano	ASJ Spring Meeting, Tamagawa Univ., (1988-03)	9
Hanazawa -88-06	HMM音韻認識におけるモデル学習の諸検討 (Studies for HMM Phoneme Recognition)	T. Hanazawa T. Kawabata K. Shikano	IEICE Technical Report SP88-22, Sendai, (1988-06)	11
Hanazawa -88-10a	Duration Control Methods for HMM Phoneme Recognition	T. Hanazawa T. Kawabata K. Shikano	ATR Technical Report, TR-1-0050, (1988-10)	18
Hanazawa -88-10b	HMM音韻認識におけるセパレートベクトル量子化の検討 (Study of Separate Vector Quantization for Phoneme Recognition)	T. Hanazawa T. Kawabata K. Shikano	ASJ Fall Meeting, Hakata, (1988-10)	28
Hanazawa -88-11	Duration Control Methods for HMM Phoneme Recognition	T. Hanazawa T. Kawabata K. Shikano	ASA-ASJ Joint Meeting, Honolulu (1988-11) (Handout)	18
Hatazaki -88-01	スペクトログラム・リーディング知識を用いた音韻認識エキスパートシステム (Phoneme Recognition Experts System Using Spectrogram Reading Knowledge)	K. Hatazaki S. Tamura T. Kawabata K. Shikano	IEICE Technical Report SP87-117 (1988-01)	30
Hatazaki -88-03	スペクトログラム・リーディング知識を用いた音韻セグメンテーションの試み (Phoneme Segmentation Using Spectrogram Reading Knowledge)	K. Hatazaki S. Tamura T. Kawabata K. Shikano	ASJ Spring Meeting, Tamagawa Univ., (1988-03)	36

Technical Publication List at the Speech Processing Department
(November, 1987 through December, 1988)

(Bold letter titles show papers written in English or with an English abstract)

Ref.ID	Title(題名)	Authors (著者)	Conference/Journal (発表先)	Page
Hatazaki -88-08	Phoneme Segmentation by an Expert System Based on Spectrogram Reading Knowledge	K. Hatazaki S. Tamura T. Kawabata K. Shikano	Seventh FASE Symposium Edinburgh, (1988-08)	38
Hatazaki -88-10	スペクトログラム・リーディング知識に基づく無声摩擦音の検出 (Segmentation of Unvoiced Fricatives Using Spectrogram Reading Knowledge)	K. Hatazaki S. Tamura T. Kawabata K. Shikano	ASJ Fall Meeting, Hakata, (1988-10)	46
Iso -88-03	音声データベース用文セットの設計 (Design of a Japanese Sentence List for a Speech Database)	*K. Iso *T. Watanabe H. Kuwabara	ASJ Spring Meeting, Tamagawa Univ., (1988-03)	48
K. Abe -88-03	入力音韻系列に応じた音声素片の選択法 (A Synthesis Unit Selection Method Adapting to an Input Phoneme)	K. Abe Y. Sagisaka	ASJ Spring Meeting, Tamagawa Univ., (1988-03)	50
K. Abe -88-11	On the Unit Selection Measure for Speech Synthesis by Rule Using Multiple Synthesis Units	K. Abe Y. Sagisaka	ASA-ASJ Joint Meeting, Honolulu (1988-11) (Handout)	52
K. Abe -88-12	On the Unit Selection Measure for Speech Synthesis by Rule Using Multiple Synthesis Units	K. Abe Y. Sagisaka	ATR Technical Report, TR-1-0053, (1988-12)	52
Katagiri -88-01	スペクトログラムを用いた音声ラベリング (Speech Labeling Using a Spectrogram)	*S. Katagiri K. Takeda Y. Sagisaka	IEICE Technical Report SP87-115, (1988-01)	62
Kawabata -87-12	日本語スペクトログラム特徴の英語との比較 (Sound Spectrogram Features in Japanese)	T. Kawabata S. Tamura K. Hatazaki	IEICE Technical Report SP87-95, (1987-12)	70
Kawabata -88-03	HMM音韻認識に基づくワードスポッティング (Word Spotting Method Based on HMM Phoneme Recognition)	T. Kawabata T. Hanazawa K. Shikano	ASJ Spring Meeting, Tamagawa Univ., (1988-03)	76
Kawabata -88-06	HMM音韻認識に基づくワードスポッティング (Word Spotting Method Based on HMM Phoneme Recognition)	T. Kawabata T. Hanazawa K. Shikano	IEICE Technical Report SP88-23, Sendai, (1988-06)	78
Kawabata -88-10	HMM音韻認識を用いた日本語名詞節の認識 (Japanese Phrase Recognition on HMM Phone Units)	T. Kawabata K. Shikano	ASJ Fall Meeting, Hakata, (1988-10)	84
Kawabata -88-11	Word Spotting Method Based on HMM Phoneme Recognition	T. Kawabata T. Hanazawa K. Shikano	ASA-ASJ Joint Meeting, Honolulu (1988-11) (Handout)	86
Kawabata -88-12	Word Spotting Method Based on HMM Phoneme Recognition	T. Kawabata T. Hanazawa K. Shikano	ATR Technical Report, TR-1-0057, (1988-12)	—

Technical Publication List at the Speech Processing Department
 (November, 1987 through December, 1988)

(Bold letter titles show papers written in English or with an English abstract)

Ref.ID	Title(題名)	Authors (著者)	Conference /Journal (発表先)	Page
Kita -88-10a	HMM音韻認識とLRパーザを用いた文節認識 (HMM Continuous Speech Recognition Using Predictive LR Parser)	K.Kita T.Kawabata H.Saito	ASJ Fall Meeting, Hakata, (1988-10)	97
Kita -88-10b	HMM音韻認識と予測LRパーザを用いた文節認識 (HMM Continuous Speech Recognition Using Predictive LR Parsing)	K.Kita T.Kawabata H.Saito	IEICE Technical Report Tokyo, (1988-10)	99
Kure matsu -87-11	音声の知的符号化-認識/合成による音声知的通信の諸問題 (Study Items for Intelligent Speech Communication with Recognition and Synthesis)	A.Kurematsu	IEICE Fall Meeting, Panel Discussion Kumamoto Univ. (1987-11)	106
Kure matsu -88-01	Prospect of a Basic Study for the Automatic Telephone Interpretation	A.Kurematsu	1st International Symposium on Advanced Man-Machine Interface Through Spoken Language, Tokyo, (1988-01)	107
Kure matsu -88-05	スペクトログラムに基づく音声表記 (Phonemic Description Based on Spectrogram)	A.Kurematsu	Science Promotion Committee, Language & Speech Group. 152 WG, (1988-05) (Handout)	111
Kure matsu -88-08	自動翻訳電話の研究動向 (Perspective of Automatic Interpreting Telephony Research)	A.Kurematsu	Journal of IEICE, vol.71 no.8, (1988-08)	120
Kure matsu -88-11	Overview of Telephone Interpretation Research at ATR	A.Kurematsu K.Shikano	ASA-ASJ Joint Meeting, Honolulu, Special Symposium, (1988-11)	123
Kuwabara -88-02	Quality Control of Speech by Modifying Formant Frequencies and Bandwidths	H.Kuwabara	ATR Technical Report TR-1-0023 (1988-02)	—
Kuwabara -88-05a	Analysis and Prediction of Vowel-Devocalization in Isolated Japanese Words	H.Kuwabara K.Takeda	ATR Technical Report TR-1-0026, (1988-05) (Kuwabara-88-05)	132
Kuwabara -88-05b	Analysis and Prediction of Vowel-Devocalization in Isolated Japanese Words	H.Kuwabara K.Takeda	Acoustical Society of America, Spring Meeting, Seattle, (1988-05) (handout)	132
Kuwabara -88-11	Speech Synthesis, Voice Conversion and Noise Reduction Research at ATR	H.Kuwabara Y.Sagisaka S.Tamura M.Miyatake K.Abe, M.Abe K.Takeda K.Shikano A.Kurematsu	2nd Symposium on Advanced Man-Machine Interface through Spoken Language, Makaha, Hawaii, (1988-11)	146

 Technical Publication List at the Speech Processing Department
 (November, 1987 through December, 1988)

(Bold letter titles show papers written in English or with an English abstract)

Ref.ID	Title(題名)	Authors (著者)	Conference /Journal (発表先)	Page
M.Abe -88-02	Voice Conversion through Vector Quantization (ベクトル量子化による声質変換)	M.Abe S.Nakamura K.Shikano H.Kuwabara	IEICE Technical Report SP87-124 (1988-02)	155
M.Abe -88-03	ベクトル量子化による声質変換の評価 (Evaluation of Voice Conversion through Vector Quantization)	M.Abe S.Nakamura K.Shikano H.Kuwabara	ASJ Spring Meeting, Tamagawa Univ., (1988-03)	161
M.Abe -88-04	Voice Conversion through Vector Quantization	M.Abe S.Nakamura K.Shikano H.Kuwabara	ICASSP'88 S14.1 New York (1988-04)	163
M.Abe -88-09	FFTスペクトルからの Signal Reconstruction による音声変換手法 (A Speech Modification Method by Signal Reconstruction Using Short-Time Fourier Transform)	M.Abe S.Tamura H.Kuwabara	IEICE Technical Report Nagoya, SP88-48, (1988-09)	167
M.Abe -88-10	FFTスペクトルからの Signal Reconstruction による音声変換手法 (Speech Modification Method by Signal Reconstruction Using Short-Time Fourier Transform)	M.Abe S.Tamura H.Kuwabara	ASJ Fall Meeting, Hakata, (1988-10)	175
M.Naka mura -88-03	コネクショニストモデルによる単語列予測の検討 (A Study of Word Category Prediction by Connectionist Models)	M.Nakamura K.Shikano	ASJ Spring Meeting, Tamagawa Univ., (1988-03)	177
M.Naka mura -88-06	英文テキストデータからのニューラルネットワークによる単語列予測モデルの検討 (A Study of English Word Category Prediction Based on Neural Networks)	M.Nakamura K.Shikano	IEICE Technical Report SP88-26, Sendai, (1988-06)	179
M.Naka mura -88-10	ニューラルネットワークによるN-gram単語列予測モデルの検討 (A Study of N-Gram Word Category Prediction Based on Neural Networks)	M.Nakamura K.Shikano	ASJ Fall Meeting, Hakata, (1988-10)	187
M.Naka mura -88-11	A Study of English Word Category Prediction Based on Neural Networks	M.Nakamura K.Shikano	ASA-ASJ Joint Meeting, Honolulu (1988-11) (Handout)	189
M.Naka mura -88-12	A Study of English Word Category Prediction Based on Neural Networks	M.Nakamura K.Shikano	ATR Technical Report, TR-1-0052, (1988-12)	189
Maruyama -87-12a	音声信号処理の基本操作を提供する機能モジュール群の作成 (Function Modules for Speech Processing)	K.Maruyama T.Kawabata	ATR Technical Report TR-1-0012 (1987-12)	—

Technical Publication List at the Speech Processing Department
 (November, 1987 through December, 1988)

(Bold letter titles show papers written in English or with an English abstract)

Ref.ID	Title(題名)	Authors (著者)	Conference/Journal (発表先)	Page
Maruyama -87-12b	対話型操作環境をワークベンチごとに設定 できるメニューシステムの作成 (User Oriented Interactive Menu System for Speech Workbench)	K.Maruyama T.Kawabata	ATR Technical Report TR-I-0011 (1987-12)	—
Maruyama -88-03	Xウィンドウ上に実現した音声処理ワーク ベンチ(Speech Processing Workbench on X- Window)	K.Maruyama T.Kawabata	ASJ Spring Meeting, Tamagawa Univ., (1988-03)	200
Maruyama -88-01	音声処理ワークベンチ (Speech Workbench)	K.Maruyama T.Kawabata	ATR Technical Report TR-I-0020, (1988-01)	—
Maruyama -88-02	高精度スペクトログラム表示ルーチン (High Quality Spectrogram Display Routine)	K.Maruyama K.Murayama T.Kawabata	ATR Technical Report TR-I-0021 (1988-02)	—
Maruyama -88-11a	Hidden Markov Modelを用いた英単語認識 (English Word Recognition Based on HMM Phoneme Models)	K.Maruyama T.Hanazawa T.Kawabata K.Shikano	ATR Technical Report, TR-I-0047, (1988-11)	—
Maruyama -88-11b	英語データベースの作成 (ATR English Database)	K.Maruyama K.Shikano	ATR Technical Report, TR-I-0048, (1988-11)	—
Miyatake -88-03	種々の発声様式における基本周波数バタンの 制御モデルの検討(Fundamental Frequency Control for Various Speaking Styles)	M.Miyatake Y.Sagisaka	ASJ Spring Meeting, Tamagawa Univ., (1988-03)	202
Miyatake -88-04	種々の発声様式における音調バタンの性質 とその制御について (Relation between Prosodic Patterns and Various Speaking Styles)	M.Miyatake Y.Sagisaka	Kinki Speech & Lan- guage Study Group, Osaka Shoin Univ. (1988-04), (Handout)	204
Miyatake -88-05a	Prosodic Characteristics and Their Control in Japanese Speech under Varying Speaking Styles	M.Miyatake Y.Sagisaka	Acoustical Society of America, Spring Meeting, Seattle, (1988-05) (handout)	214
Miyatake -88-05b	種々の発声様式における韻律パラメータの 性質について (Relation between Prosodic Patterns and Various Speaking Styles)	M.Miyatake Y.Sagisaka	ATR Technical Report TR-I-0024, (1988-05)	—
Miyatake -88-06	Prosodic Characteristics and their Control in Japanese Speech with Various Speaking Styles	M.Miyatake Y.Sagisaka	ATR Technical Report TR-I-0025, (1988-06)	214
Murayama -88-10	音声処理ワークベンチ Ver.3 (Speech Workbench, Version 3)	K.Murayama T.Kawabata	ATR Technical Report, TR-I-0037, (1988-10)	—
Niimi -88-10	タスクの複雑さを評価する尺度 (Indices to Measure the Complexity of Speech Recognition Task)	*Y.Niimi *S.Nakagawa K.Shikano T.Kawabata, etc.	Meeting on Advanced Man-Machine Interface through Spoken Lan- guage, Tokyo, (1988-10)	225

 Technical Publication List at the Speech Processing Department
 (November, 1987 through December, 1988)

(Bold letter titles show papers written in English or with an English abstract)

Ref.ID	Title(題名)	Authors (著者)	Conference/Journal (発表先)	Page
S.Naka mura -87-11	ベクトル量子化を用いたスペクトログラム の正規化 (Spectrogram Normalization Based on Vector Quantization)	S.Nakamura K.Shikano	IEICE Fall Meeting, Symposium Kumamoto Univ. (1987-11)	233
S.Naka mura -88-02	ファジィベクトル量子化を用いたスペクト ログラム正規化の検討 (Spectrogram Normalization Using Fuzzy Vector Quantization)	S.Nakamura K.Shikano	IEICE Technical Report SP87-123 (1988-02)	235
S.Naka mura -88-03	ファジィベクトル量子化を用いたスペクト ログラム正規化(Spectrogram Normalization Using Fuzzy Vector Quantization)	S.Nakamura K.Shikano	ASJ Spring Meeting, Tamagawa Univ., (1988-03)	241
S.Naka mura -88-08a	Spectrogram Normalization Using Separate Vector Quantization	S.Nakamura K.Shikano	Seventh FASE Symposium Edinburgh, (1988-08)	243
S.Naka mura -88-08b	セパレートベクトル量子化を用いたスペク トログラムの正規化 (Spectrogram Normalization Based on Separate Vector Quantization)	S.Nakamura K.Shikano	Journal of Acoustic Society of Japan, (1988-08)	251
S.Naka mura -88-10	ファジィベクトル量子化に基づく話者適応 化のHMM音素認識による評価 (Phoneme Recognition Evaluation of HMM Speaker Adaptation Using Fuzzy Vector Quantization)	S.Nakamura T.Hanazawa K.Shikano	ASJ Fall Meeting, Hakata, (1988-10)	259
S.Naka mura -88-12	ベクトル量子化話者適応化アルゴリズムの HMM音韻認識による評価(Phoneme Recognition Evaluation of HMM Speaker Adaptation Based on Vector Quantization)	S.Nakamura T.Hanazawa K.Shikano	IEICE Technical Report, SP88-105 Yokosuka, (1988-12)	261
Sagisaka -87-11	音声合成単位構成を目的とした日本語音 韻連続特性の分析(Analysis on Japanese Phonetic Sequence Characteristics for Speech Synthesis Unit Selection)	Y.Sagisaka	IEICE Fall Meeting, Kumamoto Univ. (1987-11)	269
Sagisaka -88-03	種々の音韻連続単位を用いた日本語音声合 成 (Speech Synthesis of Japanese Using Non-Uniform Phoneme Sequence Units)	Y.Sagisaka	IEICE Technical Report SP87-136, (1988-03)	270
Sagisaka -88-04	Speech Synthesis by Rule Using an Optimal Selection of Non- Uniform Synthesis Units	Y.Sagisaka	ICASSP'88 S14.8 New York, (1988-04)	276
Sagisaka -88-07	音声合成のための言語処理と韻律制御 (Language Processing and Prosody Control for Speech Synthesis)	Y.Sagisaka	Journal of Japanese Society for Artificial Intelligence, vol.3 no.4, (1988-07)	280

Technical Publication List at the Speech Processing Department
 (November, 1987 through December, 1988)

(Bold letter titles show papers written in English or with an English abstract)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal(発表先)	Page
Saito-88-06	Understanding Noisy Sentences by an LR Parser (LRパーザによる誤りを含む文の認識)	H.Saito *M.Tomita	IEICE Technical Report SP88-28, Sendai, (1988-06)	287
Saito-88-07	A Phoneme Lattice Parsing for Continuous Speech Recognition	H.Saito	ATR Technical Report TR-I-0033, (1988-07)	—
Saito-88-08	Parsing Noisy Sentences	H.Saito *M.Tomita	COLING'88 Budapest, (1988-08)	294
Sawai-88-10	時間遅れ神経回路網による音節スポットティングの検討 (A Preliminary Study on Spotting Japanese CV-Syllables by Time-Delay Neural Networks)	H.Sawai A.Waibel K.Shikano	ASJ Fall Meeting, Hakata, (1988-10)	300
Sawai-88-12	Phoneme Recognition by Scaling up Modular Time-Delay Neural Networks (モジュール構成ニューラルネットワークのスケールアップによる音韻認識)	H.Sawai A.Waibel M.Miyatake K.Shikano	IEICE Technical Report, SP88-105 Yokosuka, (1988-12)	302
Sego-87-12	Voice Conversion by Analysis-Synthesis Method	H.Segot H.Kuwabara	ATR Technical Report TR-I-0013, (1987-12)	—
Shikano-87-11	音声認識へのベクトル量子化の適用 (Vector Quantization Applied to Speech Recognition)	K.Shikano	IEICE Fall Meeting, Symposium Kumamoto Univ. (1987-11)	310
Shikano-88-01	音声認識と言語情報 (Speech Recognition and Language Knowledges)	K.Shikano	Natural Language Symposium of Information Processing Society, Tokyo, (1988-01)	312
Shikano-88-09a	コネクショニストモデルと音声認識理解 (Connectionist Models Applied to Speech Recognition)	K.Shikano	IEICE Fall Meeting, Panel, Tokyo, (1988-09)	327
Shikano-88-09b	ニューラルネットワークの音声情報処理への応用 (Neural Networks Applied to Speech Processing)	K.Shikano M.Nakamura S.Tamura A.Waibel P.Haffner H.Sawai, etc.	IEICE Special Seminar on Neural Networks, Osaka, (1988-09)	329
Shikano-88-09c	音声理解のニューパラダイム (A New Paradigm in Speech Recognition Research)	K.Shikano	AVIRG-SMG Summer Seminar, (1988-09)	336
Shikano-88-10a	音声認識におけるHMM/ニューラルネットワーク/特徴ベース (HMM, Neural Network, and Feature-Based Approaches)	K.Shikano	Joint Meeting of Electric & Information Related Societies, Symposium, Niigata Univ. (1988-10)	343

 Technical Publication List at the Speech Processing Department
 (November, 1987 through December, 1988)

(Bold letter titles show papers written in English or with an English abstract)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal(発表先)	Page
Shikano-88-10b	ニューラルネットワークの音声情報処理への応用 (Neural Network Application to Speech Processing)	K.Shikano M.Nakamura S.Tamura A.Waibel	Journal of Acoustic Society of Japan, Vol.44-10, (1988-10)	347
Shikano-88-11a	Speech Recognition Research at ATR	K.Shikano T.Kawabata A.Waibel K.Hatazaki H.Sawai S.Nakamura T.Hanazawa K.Kita A.Kurematsu	2nd Symposium on Advanced Man-Machine Interface through Spoken Language, Makaha, Hawaii, (1988-11)	361
Shikano-88-11b	音声情報の認識と理解 (Recent Trends on Speech Recognition)	K.Shikano T.Kawabata K.Hatazaki	Journal of IEICE, Special Issues for Pattern Recognition/Understanding, (1988-11)	370
Shikano-88-12	ニューラルネットワークの音声情報処理への応用	K.Shikano *S.Katagiri M.Nakamura S.Tamura A.Waibel H.Sawai P.Haffner, etc.	ATR Technical Report TR-I-0063 (1988-12)	—
Speech-87-12	Research Activities of Speech Processing Department (April 1986 through October 1987)	Speech Processing Department	ATR Technical Report TR-I-0010 (1987-12)	—
Speech-88-12	Speech Research at ATR Interpreting Telephony Research Laboratories	Speech Processing Department	ATR Technical Report TR-I-0061 (1988-12)	—
Takeda-87-12	簡易検索言語を持つ音声データベース管理システム (Speech Database Retrieval System -EAL-)	K.Takeda	ATR Technical Report TR-I-0009 (1987-12)	—
Takeda-88-01a	音声データベース管理システムの構築 (Construction of a Speech Database Management System)	K.Takeda H.Kuwabara *S.Morikawa	IEICE Technical Report SP87-116 (1988-01)	382
Takeda-88-01b	音声データベース構築のための視察による音韻ラベリング (Phoneme Labeling by Visual Examination for Speech Database Construction)	K.Takeda Y.Sagisaka H.Kuwabara	ATR Technical Report TR-I-0019 (1988-01)	—
Takeda-88-03	文発声と単語発声における音韻継続長の分析 (Vowel Duration in Isolated and Continuous Utterances)	K.Takeda Y.Sagisaka H.Kuwabara	ASJ Spring Meeting, Tamagawa Univ., (1988-03)	388

Technical Publication List at the Speech Processing Department
 (November, 1987 through December, 1988)

(Bold letter titles show papers written in English or with an English abstract)

Ref.ID	Title(題名)	Authors (著者)	Conference/Journal (発表先)	Page
Takeda -88-05	研究用日本語音声データベース利用解説書 (User's Manual for Japanese Speech Database)	K.Takeda Y.Sagisaka *S.Katagiri M.Abe H.Kuwabara	ATR Technical Report TR-I-0026, (1988-05)	—
Takeda -88-10a	入力音韻系列に応じた音声合成素片選択法の改良 (An Improvement of the Optimal Units Selection Method for Speech Synthesis)	K.Takeda Y.Sagisaka K.Abe H.Kuwabara	ASJ Fall Meeting, Hakata, (1988-10)	390
Takeda -88-10b	On Sentence Level Factors Governing Segmental Duration in Japanese	K.Takeda Y.Sagisaka H.Kuwabara	ATR Technical Report, TR-I-0051, (1988-10)	392
Takeda -88-10c	研究用日本語音声データベースの構築 (A Japanese Speech Database for Various Kinds of Research Purposes)	K.Takeda Y.Sagisaka *S.Katagiri H.Kuwabara	Journal of Acoustic Society of Japan, (1988-10)	413
Takeda -88-11	On Sentential Effects in Segmental Duration Control of Japanese	K.Takeda Y.Sagisaka H.Kuwabara	ASA-ASJ Joint Meeting, Honolulu (1988-11) (Handout)	392
Tamura -88-01	Noise Reduction Using Neural Networks (Neural Networkを使った波形入出力による雑音抑圧)	S.Tamura A.Waibel	IEICE Technical Report SP87-112, (1988-01)	421
Tamura -88-03	Noise Reduction through Waveform Input and Output Using Neural Networks (Neural Networkを使った波形入出力による雑音抑圧)	S.Tamura A.Waibel	ASJ Spring Meeting, Tamagawa Univ., (1988-03)	426
Tamura -88-04	Noise Reduction Using Connectionist Models	S.Tamura A.Waibel	ICASSP'88 S12.7 New York, (1988-04)	428
Tamura -88-10	波形入出力による雑音抑圧ニューラルネットワークの解析 (An Analysis of a Noise Reduction Neural Network Which Takes Waveforms as Input and Output)	S.Tamura	ASJ Fall Meeting, Hakata, (1988-10)	432
Waibel -87-11	Phoneme Recognition Using Time-Delay Neural Networks	A. Waibel T.Hanazawa *G. Hinton K.Shikano *K. Lang	ATR Technical Report TR-I-0006 (1987-11)	—
Waibel -87-12	Phoneme Recognition Using Time-Delay Neural Networks (時間遅れ神経回路網(TDNN)による音韻認識)	A.Waibel	IEICE Technical Report SP87-100 (1987-12)	434

 Technical Publication List at the Speech Processing Department
 (November, 1987 through December, 1988)

(Bold letter titles show papers written in English or with an English abstract)

Ref.ID	Title(題名)	Authors (著者)	Conference/Journal (発表先)	Page
Waibel -88-04a	Speech Recognition Using Time-Delay Neural Networks	A.Waibel	Snowbird Conference Salt Lake City (1988-04)	440
Waibel -88-04b	Phoneme Recognition: Neural Networks vs. Hidden Markov Models	A.Waibel K.Shikano T.Hanazawa *G. Hinton *K. Lang	ICASSP'88 S3.3 New York (1988-04)	441
Waibel -88-05	Speech Recognition Using Time-Delay Neural Networks	A.Waibel	Acoustical Society of America, Spring Meeting, Seattle, (1988-05)	445
Waibel -88-08	Modularity and Scaling in Large Phonemic Neural Networks	A. Waibel H.Sawai K.Shikano	ATR Technical Report TR-I-0033, (1988-08)	—
Waibel -88-10	Phoneme Recognition by Modular Construction of Time-Delay Neural Networks	A.Waibel H.Sawai K.Shikano	ASJ Fall Meeting, Hakata, (1988-10)	446
Waibel -88-11a	Modularity in Neural Networks for Speech Recognition	A.Waibel	IEEE Conf. on Neural Information Processing Systems, Denver, U.S.A., (1988-11)	447
Waibel -88-11b	Incremental Learning of Large Phonetic Neural Networks from Smaller Subsets	A.Waibel	USA-Japan Joint Acoustical Society Meeting, Honolulu, U.S.A., (1988-11)	448
Yamazaki -88-09	種々の合成素片を用いる規則合成システムにおける韻律制御法の改良 (Improvement for Prosodic Parameter Control in Speech Synthesis-by-Rule System)	*T.Yamazaki K.Takeda H.Kuwabara	ATR Technical Report, TR-I-0040, (1988-09)	—

Technical Publication List at the Speech Processing Department
(January, 1989 through October, 1989)
(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
Cheveigné -89-10a	Narrowed Autocoincidence of Nerve Spike Patterns and Pure Tone Pitch	A. Cheveigné H. Kuwabara	ASJ Fall meeting, 1-2-4, Toyama, (1989-10)	1
Cheveigné -89-10b	The MapSignal Remote Speech Editor	A. Cheveigné	ASJ Fall meeting, 3-4-15, Toyama, (1989-10)	3
Cheveigné -89-10c	Pitch, and the Narrowed Autocoincidence Histogram	A. Cheveigné	1st International Conference on Music Perception and Cognition (ICMPC), Kyoto,(1989-10)	5
Dang -89-10	Deterministic Boltzmann Machines for Phoneme Recognition	J-C. Dang H. Sawai	ASJ Fall meeting, 1-1-23, Toyama, (1989-10)	9
Endo -89-08	ニューラルネットワークによる予測モデルを用いた音韻認識 (Phoneme Recognition Experiment by Neural Prediction Models)	T. Endo (Waseda Univ.) S. Tamura M. Nakamura	ATR Technical Report TR-I-0107, (1989-08)	—
Haffner -89-03a	Fast Back-Propagation Learning Methods for Large Phonemic Neural Networks	P. Haffner H. Sawai A. Waibel K. Shikano	ASJ Spring meeting, 1-6-14, Waseda, (1989-03)	11
Haffner -89-03b	Fast Back-Propagation Learning Methods for Neural Networks in Speech Recognition	P. Haffner H. Sawai A. Waibel K. Shikano	IEICE Spring meeting, 5A-1-1, Kinki Univ., (1989-03)	13
Haffner -89-09	Fast Back-Propagation Learning Methods for Large Phonemic Neural Networks	P. Haffner H. Sawai A. Waibel K. Shikano	European Conference on Speech Communica- tion and Technology, pp553-556, Paris, (1989-09)	15
Hanazawa -89-03	HMM音韻モデル文節認識による評価 (Evaluation of HMM Phone Units through Japanese Phrase Recognition)	T. Hanazawa K. Kita T. Kawabata K. Shikano	ASJ Spring meeting, 3-6-6, Waseda, (1989-03)	19
Hanazawa -89-10a	ベクトル量子化話者適応アルゴリズムのHMM文節認識による評価 (Evaluation of VQ Based Speaker Adaptation Algorithm through HMM Japanese Phrase Recognition)	T. Hanazawa S. Nakamura T. Kawabata K. Shikano	ASJ Fall meeting, 2-P-(18), Toyama, (1989-10)	21
Hanazawa -89-10b	Hidden Markov モデルによる日本語有声破裂音の認識 (Recognition of Japanese Voiced Stops Using Hidden Markov Models)	T. Hanazawa T. Kawabata K. Shikano	Journal of ASJ, Vol.10, No. 10, pp776-785 (1989-10)	23

Technical Publication List at the Speech Processing Department
(January, 1989 through October, 1989)
(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
Hatazaki -89-03a	スペクトログラム・リーディング知識による音韻セグメンテーションの評価 (Evaluation of Phoneme Segmentation Using Spectrogram Reading Knowledge)	K. Hatazaki Y. Komori	ASJ Spring meeting, 2-P-(2), Waseda, (1989-03)	33
Hatazaki -89-03b	スペクトログラム・リーディング知識を用いた音韻セグメンテーション・エキスパートシステム (Phoneme Segmentation Expert System Using Spectrogram Reading Knowledge)	K. Hatazaki Y. Komori	ATR Technical Report TR-I-0072, (1989-03)	—
Hatazaki -89-03c	スペクトログラム・リーディング知識に基づく音韻セグメンテーション知識 (Phoneme Segmentation Knowledge Based on Spectrogram Reading)	K. Hatazaki Y. Komori	ATR Technical Report TR-I-0073, (1989-03)	—
Hatazaki -89-05	Phoneme Segmentation Using Spectrogram Reading Knowledge	K. Hatazaki Y. Komori T. Kawabata K. Shikano	ICASSP'89, 5.8.2, pp393-396 Glasgow, (1989-05)	35
Hirata -89-03	HMM-LR法を用いた文節認識における継続時間制御パラメータ変換法の検討 (Study on Duration Control for HMM-LR Continuous Speech Recognition)	Y. Hirata (Toyoashi Tech.) T. Kawabata T. Hanazawa	ATR Technical Report TR-I-0076, (1989-03)	—
K. Abe -89-03	波形重ね合わせ法による合成音の品質について (Quality Evaluation for Synthesized Speech Using Wave Overlap Adding)	K. Abe Y. Sagisaka H. Kuwabara	ASJ Spring meeting, 1-7-16, Waseda, (1989-03)	39
K. Abe -89-10	音声合成素片の接続方法の検討 (A Study on Speech Synthesis Units Concatenation)	K. Abe Y. Sagisaka K. Takeda	ASJ Fall meeting, 3-P-5, Toyama, (1989-10)	41
Kawabata -89-03	音韻パレキシティの提案 (Task Entropy and Phone Perplexity)	T. Kawabata K. Shikano K. Kita	ASJ Spring meeting, 3-6-12, Waseda, (1989-03)	43
Kawabata -89-05	Island-Driven Continuous Speech Recognizer Using Phoneme-Based HMM Word Spotting	T. Kawabata K. Shikano	ICASSP'89, 59.7, pp461-464, Glasgow, (1989-05)	45
Kawabata -89-10	構成的ニューラルネットによる音声認識 (Constructive Neural Network for Speech Recognition)	T. Kawabata	ASJ Fall meeting, 1-1-26, Toyama, (1989-10)	49
Kita -89-03	HMM-LR 連続音声認識システムにおける計算量削減のための一検討 (Computing Amount Reduction in HMM-LR Continuous Speech Recognition System)	K. Kita T. Kawabata T. Morimoto	ASJ Spring meeting, 3-6-4, Waseda, (1989-03)	51

Technical Publication List at the Speech Processing Department
(January, 1989 through October, 1989)

(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
Kita -89-05	HMM Continuous Speech Recognition Using Predictive LR Parsing	K. Kita T. Kawabata H. Saito	ICASSP'89, S13.3, pp703-706, Glasgow, (1989-05)	53
Kita -89-06	HMM音韻認識と拡張LR構文解析法を用いた連続音韻認識 (HMM Continuous Speech Recognition Using Generalized LR Parsing)	K. Kita T. Kawabata H. Saito	ATR Technical Report TR-I-0082, (1989-06)	—
Kita -89-08	Parsing Continuous Speech by HMM-LR Method	K. Kita T. Kawabata H. Saito	International Workshop on Parsing Technologies, pp126-131 Pittsburgh, (1989-08)	57
Kita -89-10	SL-TRANSにおける文節音声識 (Speech Recognition Method in SL-TRANS)	K. Kita T. Sakano J. Hosaka T. Kawabata	Information Processing Society of Japan, Fall meeting, Kyusyu Institute of Technology, (1989-10)	63
Komori -89-03	スペクトログラム,リーディング知識に基づく音韻認識エキスパートシステムの構築 (Phoneme Recognition Expert System Using Spectrogram Reading Knowledge)	Y. Komori K. Hatazaki T. Tanaka T. Kawabata	ASJ Spring meeting, 3-6-11, Waseda, (1989-03)	65
Komori -89-06	スペクトログラム,リーディング知識とニューラルネットワークを用いた音韻認識エキスパートシステム (Phoneme Recognition Expert Spectrogram Reading Knowledge and Neural Networks)	Y. Komori K. Hatazaki T. Tanaka T. Kawabata K. Shikano	IEICE Technical Report, SP89-33, Iwate Univ, (1989-06)	67
Komori -89-09	Phoneme Recognition Expert System Using Spectrogram Reading Knowledge and Neural Networks	Y. Komori K. Hatazaki T. Tanaka T. Kawabata K. Shikano	European Conference on Speech Communication and Technology, pp549-552, Paris, (1989-09)	75
Komori -89-10	スペクトログラム,リーディング知識に基づく音韻認識エキスパートシステムにおける音韻識別ニューラルネットワークの融合法の検討 (Combining Phoneme Identification Neural Networks into an Expert System Using Spectrogram Reading Knowledge)	Y. Komori K. Hatazaki T. Tanaka T. Kawabata K. Shikano	ASJ Fall meeting, 3-1-14, Toyama, (1989-10)	79
Kurematsu -89-09	ATR Japanese Speech Database as a Tool of Speech Recognition and Synthesis	A. Kurematsu K. Takeda H. Kuwabara K. Shikano	ESCA-Workshop on Speech I/O Assessment, 2.3.1, Netherlands, (1989-09)	81

Technical Publication List at the Speech Processing Department
(January, 1989 through October, 1989)

(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
Kuwabara -89-05	Construction of a Large-Scale Japanese Speech Database and its Management System	H. Kuwabara K. Takeda Y. Sagisaka S. Morikawa(TIS) T. Watanabe(NEC)	ICASSP'89 S10b.12, pp560-563, Glasgow, (1989-05)	85
Kuwabara -89-07a	日本語促音の特徴 (Characteristics of a Japanese Phoneme "Sokuon")	H. Kuwabara N. Yoshida (Kyoto Univ.)	ATR Technical Report TR-I-0087, (1989-07)	—
Kuwabara -89-07b	日本語におけるアクセントの機能と音韻句の問題 (Accentuation of Japanese Language and Problems of Phonological Phrase)	H. Kuwabara Y. Iwai (Kyoto Univ.)	ATR Technical Report TR-I-0088, (1989-07)	—
Kuwabara -89-07c	研究用ATR音声データベースの作成 (Construction of ATR Japanese Speech Database as a Research Tool)	H. Kuwabara M. Abe Y. Sagisaka K. Takeda	ATR Technical Report TR-I-0086, (1989-07)	—
M.Abe -89-03	入力スペクトル情報の利用による声質変換法の高度化 (Improvement of Voice Conversion by Usage of Input Speech Spectrum)	M. Abe S. Nakamura K. Shikano H. Kuwabara	ASJ Spring meeting, 1-7-23, Waseda, (1989-03)	89
M.Abe -89-04a	連続音声の基本周波数データベース (Pitch Frequency Database on Continuous Speech)	M. Abe H. Kuwabara	ATR Technical Report TR-I-0078, (1989-04)	—
M.Abe -89-04b	連続音声データベースにおける言語韻律情報 (Integrating Linguistic and Prosodic Information in a Continuous Speech Database)	M. Abe Y. Sagisaka H. Kuwabara	ATR Technical Report TR-I-0079, (1989-04)	—
M.Abe -89-05	A New Speech Modification Method by Signal Reconstruction	M. Abe S. Tamura H. Kuwabara	ICASSP'89, S11.9, pp592-595 Glasgow, (1989-05)	91
M.Abe -89-08	FFTスペクトルからの信号再生法による音声変換手法 (A Speech Modification Method by Signal Reconstruction Using Short-time Fourier Transform)	M. Abe S. Tamura H. Kuwabara	Journal of IEICE (D-II), Vol.J72-D-II, No.8, pp1180-1186, (1989-08)	95
M.Abe -89-10	言語韻律情報を持つ連続音声の基本周波数データベース (The Integration of Linguistic, Prosodic Information and Fundamental Frequency in a Continuous Speech Database)	M. Abe Y. Sagisaka H. Kuwabara	ASJ Fall meeting, 2-3-22, Toyama, (1989-10)	102
M.Nakamura -89-05	English Word Category Prediction Based on Neural Networks	M. Nakamura K. Shikano	ICASSP'89, S13.10, pp731-734, Glasgow, (1989-05)	104

Technical Publication List at the Speech Processing Department
(January, 1989 through October, 1989)
(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
M.Nakamura -89-09	ニューラルネット開発用ワークベンチシステムネットワークエディタおよびモニタ機能について (Neural Net Workbench System)	M. Nakamura S. Tamura M. Miyatake H. Sawai	ATR Technical Report TR-I-0113, (1989-09)	—
M.Nakamura -89-10	ニューラルネット開発用ワークベンチシステムネットワークエディタおよびモニタ機能について (Neural Net Workbench System)	M. Nakamura S. Tamura M. Miyatake H. Sawai	ASJ Fall meeting, 2-P-(26), Toyama, (1989-10)	108
Maruyama -89-01	HMM音韻連結学習を用いた英単語音声の認識 (English Word Recognition Using HMM Phone Concatenated Training)	K. Maruyama T. Hanazawa T. Kawabata K. Shikano	IEICE Technical Report, Osaka Univ., SP88-119, (1989-01)	110
Maruyama -89-03	HMM音韻連結学習を用いた英単語音声の認識 (English Word Recognition Using HMM Phone Concatenated Training)	K. Maruyama T. Hanazawa T. Kawabata K. Shikano	ASJ Spring meeting, 1-6-22, Waseda, (1989-03)	117
Maruyama -89-09	高精度HMMを用いた英単語認識 (English Word Recognition Using Multiple Codebooks)	K. Maruyama T. Hanazawa T. Kawabata K. Shikano	ATR Technical Report TR-I-0108, (1989-09)	—
Maruyama -89-10	HMM音韻連結学習とNETgramを用いた英単語音声の認識 (English Word Recognition Using HMM Phone Concatenated Training and NETgram)	K. Maruyama M. Nakamura T. Kawabata K. Shikano	ASJ Fall meeting, 2-P-(8), Toyama, (1989-10)	119
Minami -89-07	TDNN音韻スポッティングと拡張LRパーザを用いた文節音声認識 (Continuous Speech Recognition Using TDNN Phoneme Spotting and Generalized LR Parser)	Y. Minami M. Miyatake H. Sawai K. Shikano	ATR Technical Report TR-I-0085, (1989-07)	—
Minami -89-10	TDNN音韻スポッティングと拡張LRパーザを用いた分節音声認識 (Continuous Speech Recognition Using TDNN Phoneme Spotting and Generalized LR Parser)	Y. Minami M. Miyatake H. Sawai K. Shikano	ASJ Fall meeting, 3-1-11, Toyama, (1989-10)	121
Miyatake -89-03a	時間遅れニューラルネットワークを用いた音韻スポッティング法 (Phoneme Spotting Methods Using Time- Delay Neural Networks)	M. Miyatake H. Sawai K. Shikano	IEICE Spring meeting, SA-1-4, Kinki Univ., (1989-03)	123
Miyatake -89-03b	全音韻を統合した時間遅れ神経回路網(TDNN)による音韻スポッティング (Spotting Phonemes Using Integrated Time- Delay Neural Networks (TDNN))	M. Miyatake H. Sawai K. Shikano	ASJ Spring meeting, 2-P-(24), Waseda, (1989-03)	125

Technical Publication List at the Speech Processing Department
(January, 1989 through October, 1989)
(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
Miyatake -89-06	連続音声の中のスポッティングのためのTDNN構成法 (How to Apply TDNN to Phoneme Spotting in Continuous Speech)	M. Miyatake H. Sawai K. Shikano	IEICE Technical Report, SP89-32, Iwate Univ., (1989-06)	127
Miyatake -89-08a	時間遅れ神経回路網(TDNN)による音韻スポッティングのための効果的学習法 (Effective Training Methods for Spotting Japanese Phonemes Using Time-Delay Neural Networks)	M. Miyatake H. Sawai K. Shikano	ATR Technical Report TR-I-0103, (1989-08)	—
Miyatake -89-08b	音声研究用ワークベンチの処理の自動化 (An Automatic Hardcopy Tool Based on Speech Workbench)	M. Miyatake Y. Sagisaka	ATR Technical Report TR-I-0106, (1989-08)	—
Miyatake -89-08c	会話文音声生成のための音声合成、およびニューラルネットワークの連続音声認識への適用 (Prosody Control for Conversational Speech Synthesis, and Neural Networks for Continuous Speech Recognition)	M. Miyatake	ATR Technical Report TR-I-0112, (1989-08)	—
Miyatake -89-10	時間遅れ神経回路網(TDNN)による音韻スポッティングの改良 (Improvement on Spotting Phonemes Using Time-Delay Neural Networks (TDNN))	M. Miyatake H. Sawai K. Shikano	ASJ Fall meeting, 1-1-25, Toyama, (1989-10)	133
Poser -89-01a	On Phrasal Downtrends in F0 in Japanese	W. Poser	Kinki Society for Phonetics Meeting, (1989-01)	135
Poser -89-01b	Modified MITalk	W. Poser M. Abe	ATR Technical Report TR-I-0066, (1989-01)	—
Poser -89-03a	Implementation of an F0 Model for Japanese Incorporating Downstep	W. Poser Y. Sagisaka	ASJ Spring meeting, 1-7-5, Waseda, (1989-03)	150
Poser -89-03b	Tools for Fundamental Frequency Modelling	W. Poser	ATR Technical Report TR-I-0069, (1989-03)	—
Poser -89-03c	Modelling Phrasal Level F0 Phenomena in Japanese	W. Poser Y. Sagisaka	IEICE Technical Report, SP88-160, Tokyo, (1989-03)	152
S.Nakamura -89-02	ファジィベクトル量子化によるスペクトログラムの正規化 (Spectrogram Normalization Using Fuzzy Vector Quantization)	S.Nakamura K. Shikano	Journal of ASJ, Vol.45, No.2, pp107-114, (1989-02)	158

Technical Publication List at the Speech Processing Department
(January, 1989 through October, 1989)

(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
S.Nakamura -89-03	ベクトル量子化話者適応化アルゴリズムのHMM音韻認識による評価 (Phoneme Recognition Evaluation of HMM Speaker Adaptation Using Vector Quantization)	S. Nakamura T. Hanazawa K. Shikano	ASJ Spring meeting, 1-6-23, Waseda, (1989-03)	166
S.Nakamura -89-05	Speaker Adaptation Applied to HMM and Neural Networks	S. Nakamura K. Shikano	ICASSP'89, S3.3, pp89-92, Glasgow, (1989-05)	168
S.Nakamura -89-06	時間遅れ神経回路網(TDNN)における入力パラメータの評価と話者適応化 (Comparison of Input Parameters for Time-Delay Neural Network(TDNN) and Implementation of VQ-Based Speaker Adaptation)	S. Nakamura K. Shikano	IEICE Technical Report, SP89-18, Iwate Univ., (1989-06)	172
S.Nakamura -89-08a	セパレートベクトル量子化に基づく話者適応化 (Speaker Adaptation through Separate Vector Quantization)	S. Nakamura K. Shikano	ATR Technical Report TR-I-0095, (1989-08)	-
S.Nakamura -89-08b	ファジィベクトル量子化に基づく話者適応化 (Speaker Adaptation through Fuzzy Vector Quantization)	S. Nakamura K. Shikano	ATR Technical Report TR-I-0096, (1989-08)	-
S.Nakamura -89-08c	ベクトル量子化話者適応のHMM音韻認識への適応 (VQ-Based Speaker Adaptation Applied to HMM Phoneme Recognition)	S. Nakamura T. Hanazawa K. Shikano	ATR Technical Report TR-I-0097, (1989-08)	-
S.Nakamura -89-08d	ベクトル量子化話者適応の時間遅れ神経回路網による音韻認識への適用 (VQ-Based Speaker Adaptation Applied to Time-Delay Neural Network Phoneme Recognition)	S. Nakamura K. Shikano	ATR Technical Report TR-I-0098, (1989-08)	-
S.Nakamura -89-08e	話者重畳型HMMを用いた調音様式の話者適応化 (Speaker Adaptation of Articulatory Variation by Supplemented HMM)	S. Nakamura H. Hattori K. Shikano	ATR Technical Report TR-I-0099, (1989-08)	-
S.Nakamura -89-08f	ベクトル量子化話者適応化の研究 (A Study of VQ-based Speaker Adaptation)	S. Nakamura	ATR Technical Report TR-I-0100, (1989-08)	-
S.Nakamura -89-08g	話者適応化における写像方法の比較 (A Comparative Study of Spectral Mapping Methods on Speaker Adaptation)	S. Nakamura K. Shikano	ATR Technical Report TR-I-0101, (1989-08)	-
S.Nakamura -89-10a	複数話者HMM学習を用いた話者適応化の音韻認識による評価 (A Study for Supplementation of Speaker Articulatory Variation on HMM Speaker Adaptation)	S. Nakamura H. Hattori K. Shikano	ASJ Fall meeting, 2-P-(15), Toyama, (1989-10)	180

Technical Publication List at the Speech Processing Department
(January, 1989 through October, 1989)

(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
S.Nakamura -89-10b	教師つき話者適応化における写像方法の比較 (A Comparative Study of Spectral Mapping Methods on Supervised Speaker Adaptation)	S. Nakamura K. Shikano	ASJ Fall meeting, 2-P-(14), Toyama, (1989-10)	182
Sagisaka -89-03	情報量尺度を用いた音声単位セットの構成法 (On the Design of a Speech Synthesis Unit Set Using Entropy Measure)	Y. Sagisaka	ASJ Spring meeting, 1-7-20, Waseda, (1989-03)	184
Sagisaka -89-05	The Integration of Rules and Data in a Speech Synthesis System	Y. Sagisaka	Speech Tech'89 New York, (1989-05)	186
Sagisaka -89-10	統語構造に基づくF0パターン概形の制御 (On the Control of Global F0 Pattern Using Simple Syntactic Information)	Y. Sagisaka	ASJ Fall meeting, 3-P-13, Toyama, (1989-10)	189
Sawai -89-03	Spotting Phonemes by Hierarchical Construction of Time-Delay Neural Networks	H. Sawai M. Miyatake K. Shikano	ASJ Spring meeting, 2-P-(25), Waseda, (1989-03)	191
Sawai -89-05	Spotting Japanese CV-Syllables and Phonemes Using Time-Delay Neural Networks	H. Sawai A. Waibel M. Miyatake K. Shikano	ICASSP'89 S.1.7, pp25-28, Glasgow, (1989-05)	193
Sawai -89-06	Parallelism, Hierarchy, Scaling in Time-Delay Neural Networks for Spotting Japanese Phonemes/CV-Syllables	H. Sawai A. Waibel P. Haffner M. Miyatake K. Shikano	IEEE International Joint Conference on Neural Networks, (IEEE/INNS) Washington, D.C., (1989-06)	197
Sawai -89-07a	時間遅れ神経回路網を用いた音韻/音節スポットティング (Spotting Phonemes and Syllables using Time-Delay Neural Networks)	H. Sawai M. Miyatake A. Waibel K. Shikano	ATR Technical Report TR-I-0089, (1989-07)	-
Sawai -89-07b	Parallelism, Hierarchy, Scaling in Time-Delay Neural Networks for Spotting Phonemes and CV-Syllables	H. Sawai A. Waibel P. Haffner M. Miyatake K. Shikano	ATR Technical Report TR-I-0090, (1989-07)	-
Sawai -89-08	連続音声認識のための時間遅れ神経回路網を用いた音韻/音節スポットティング (Spotting Phonemes and Syllables for Continuous Speech Recognition Using Time-Delay Neural Networks)	H. Sawai M. Miyatake A. Waibel K. Shikano	Journal of IEICE,(DII), invited paper, Vol.J72-D-II, No.8, pp1151-1158, (1989-08)	205

Technical Publication List at the Speech Processing Department
(January, 1989 through October, 1989)

(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
Seiyama -89-10	複素ケプストラムによる周期成分波形を零位相化することによる基本周波数抽出法 (A Pitch Extraction Method Using Phase Equalization of Impulse Train by Complex Cepstrum)	N.Seiyama (NHK) T.Umeda	ASJ Fall meeting, 2-3-14, Toyama, (1989-10)	213
Shikano -89-01	Research Activities of the Speech Processing Department (October 1987 through December 1988)	K. Shikano	ATR Technical Report TR-I-0065, (1989-01)	—
Shikano -89-05	ニューラルネットワークの音声情報処理への応用 (Neural Networks Applied to Speech Processing)	K. Shikano	Seminar of Information Processing Society in Kansai, pp67-79, (1989-05)	215
Shikano -89-09	音韻認識エキスパートシステム (Phoneme Recognition Expert System)	K. Shikano K. Hatazaki Y. Komori	IEICE Fall meeting, PD-1-4, Kanagawa Univ., (1989-09)	228
Takagi -89-10	声質変換による、感情表現制御のための物理的要因の解析 (Analysis of Acoustic Features for Control of Emotional Expressions of Speech by Voice-Quality Conversion)	T. Takagi (NHK) H. Kuwabara (NHK) T.Umeda	ASJ Fall meeting, 3-3-4, Toyama, (1989-10)	230
Takeda -89-05	An Approach for Normalizing Coarticulatory Variations through Expansion of Acoustic Properties into Phonetic Features	K. Takeda	ATR Technical Report TR-I-0081, (1989-05)	—
Takeda -89-09	Adaptive Manipulation of Non-Uniform Synthesis Units Using Multi-level Unit Transcription	K. Takeda K. Abe Y. Sagisaka H. Kuwabara	European Conference on Speech Communication and Technology, pp195-198, Paris, (1989-09)	232
Takeda -89-10	合成単位の選択における音韻的知識と韻律的屬性の処理の統合 (Unit Selection Measurement Combining Phonetic and Prosodic Criteria for Synthesis-by-Rule)	K. Takeda Y. Sagisaka H. Kuwabara	ASJ Fall meeting, 3-P-16, Toyama, (1989-10)	236
Tamura -89-01	An Analysis of a Noise Reduction Neural Network which Takes Waveforms as Input and Output	S. Tamura	IEICE Technical Report, SP88-124, (1989-01)	238
Tamura -89-05	An Analysis of a Noise Reduction Neural Network	S. Tamura	ICASSP'89, A. 1a.6, pp2001-2004, Glasgow, (1989-05)	246
Tamura -89-06	On Interpretations of a Feed-Forward Neural Network	S. Tamura	IEEE International Joint Conference on Neural Networks ,(IEEE/INNS), II-584, Washington, D.C., (1989-06)	250

Technical Publication List at the Speech Processing Department
(January, 1989 through October, 1989)

(Bold letter titles show papers written in English or with English abstract.)

Ref.ID	Title(題名)	Authors(著者)	Conference/Journal (発表先)	Page
Tamura -89-09	ニューラルネットワークを用いた音声処理 (Neural Networks Applied to Speech Processing)	S. Tamura H. Sawai M. Nakamura K. Shikano	The Journal of the Institute of Television Engineers of Japan, Vol.43, No.9, pp935-943, (1989-09)	258
Tanaka -89-08a	X Window System, Version II ポケットガイド-Xlib編一 (X Window System, Version II, Pocket Guide - Xlib -)	T. Tanaka	ATR Technical Report TR-I-0091, (1989-08)	—
Tanaka -89-08b	X Window System, Version II ポケットガイド-Tool Kit編一 (X Window System, Version II, Pocket Guide - Tool Kit -)	T. Tanaka	ATR Technical Report TR-I-0092, (1989-08)	—
Waibel -89-03	Phoneme Recognition Using Time-Delay Neural Networks	A. Waibel T.Hanazawa G.Hinton (Tronto Univ.) K.Shikano K.Lang (CMU)	IEEE Tr.ASSP, Vol.37, No.3, pp328-339, (1989-03)	271
Waibel -89-05	Consonant Recognition by Modular Construction of Large Phonemic Time-Delay Neural Networks	A. Waibel H. Sawai K. Shikano	ICASSP'89, S.3.9, pp112-115, Glasgow, (1989-05)	267

List of Technical Publications from Speech Processing Department
(November 1989 through December 1990)
(Bold titles indicate papers written in English)

Ref.ID	Title(題名)	Authors(著者)	Journal(掲載誌)	Page
Abe,K 89-11	音韻環境に応じた音声合成素片の接続方法の検討 [On the Concatenation of Speech Synthesis Units according to Unit Extraction Context]	安部 勝雄 武田 一哉 匂坂 芳典 [K. Abe K. Takeda Y. Sagisaka]	電子情報通信学会技術研究報告 SP89-66, pp.17-22, 1989. [IECIE Technical Report, SP89-66, pp.17-22, 1989]	1
Abe,K 90-2a	波形重ね合わせ法による合成音の品質について [Quality Evaluation for Synthesized Speech Using Wave Overlap Adding]	安部 勝雄 匂坂 芳典 森原 尚夫 [K. Abe Y. Sagisaka H. Kuwabara]	ATR Technical Report TR-I-0135	-
Abe,K 90-2b	音韻環境に応じた音声合成素片の接続方法の検討 [On the Concatenation of Speech Synthesis Units According to Unit Extraction Context]	安部 勝雄 武田 一哉 匂坂 芳典 [K. Abe K. Takeda Y. Sagisaka]	ATR Technical Report TR-I-0136	-
Abe,K 90-3	規則合成におけるパワー制御の検討 [A Study on Power Control of Speech Synthesis Units]	安部 勝雄 武田 一哉 匂坂 芳典 [K. Abe K. Takeda Y. Sagisaka]	日本音響学会平成2年度春季研究発表会講演論文集, 1-4-12, pp.201-202, 1990. [Proc. of ASJ Spring Meeting, 1-4-12, pp.201-202, 1990]	7
Abe,M 89-5	Voice Conversion Through Vector Quantization	M. Abe H. Kuwabara S. Nakamura K. Shikano	The Journal of the Acoustical Society of Japan, pp.71-76, 1989.	9
Abe,M 89-11	Fundamental Frequency Database with Linguistic and Phonetic Information	M. Abe Y. Sagisaka H. Kuwabara	The Journal of the Acoustical Society of America, Supplement 1, Vol.86, O8, pp.536, Fall, 1989.	15
Abe,M 89-12	Cross-Language Voice Conversion	M. Abe	ATR Technical Report TR-I-0126	-
Abe,M 90-2	言語間にわたる声質変換 [Cross-Language Voice Conversion]	阿部 匡伸 [M. Abe]	電子情報通信学会技術研究報告 SP89-123, pp.31-38, 1990. [IECIE Technical Report, SP89-123, pp.31-38, 1990]	16
Abe,M 90-3a	声質変換の研究 [A Study on Voice Conversion]	阿部 匡伸 [M. Abe]	ATRジャーナル, No.7, pp.8-11, 1990. [ATR Journal, No.7, pp.8-11, 1990]	24
Abe,M 90-3b	Cross-Language Voice Conversion	M. Abe	日本音響学会平成2年度春季研究発表会講演論文集, 3-4-11, pp.289-290, 1990. [Proc. of ASJ Spring Meeting, 3-4-11, pp.289-290, 1990]	28

List of Technical Publications from Speech Processing Department
(November 1989 through December 1990)
(Bold titles indicate papers written in English)

Ref.ID	Title(題名)	Authors(著者)	Journal(掲載誌)	Page
Abe,M 90-4	Cross-language Voice Conversion	M. Abe K. Shikano H. Kuwabara	Proceedings of 1990 International Conference on Acoustics, Speech, and Signal Processing, 56a.14, pp.345-348, Albuquerque, 1990.	30
Abe,M 90-6	Voice Conversion for an Interpreting Telephone	M. Abe K. Shikano H. Kuwabara	Proceedings of the Tutorial and Research Workshop on Speaker Characterization in Speech Technology, pp.40-45, Edinburgh, 1990.	34
Abe,M 90-9a	音声セグメントを変換の単位とする声質変換 [A Segment Model Based Approach to Voice Conversion]	阿部 匡伸 嵯峨山 茂樹 梅田 哲夫 [M. Abe S. Sagayama T. Umeda]	日本音響学会平成元年度秋季研究発表会講演論文集, 3-6-11, pp.287-288, 1990. [Proc. of ASJ Fall meeting, 3-6-11, pp.287-288, 1990]	40
Abe,M 90-9b	研究用日本語音声データベース利用解説書(連続音声データベース編) [Speech Database User's Manual]	阿部 匡伸 匂坂 芳典 梅田 哲夫 森原 尚夫 [M. Abe Y. Sagisaka T. Umeda H. Kuwabara]	ATR Technical Report TR-I-166	-
Abe,M 90-11	Statistical Study on Voice Individuality Conversion across Different Languages	M. Abe S. Sagayama	Proceedings of 1990 International Conference on Spoken Language Processing, 5.8, pp.157-160, Kobe, 1990.	42
Campbell 90-11a	Evidence for a Syllable-based Model of Speech Timing	Nick Campbell	Proceedings of 1990 International Conference on Spoken Language Processing, 1.3, pp.9-12, Kobe, 1990.	46
Campbell 90-11b	Duration, Pitch and Diphones in the CSTR TTS System	Nick Campbell S.D. Isard A.I.C.Monaagham J.Vechhoeveen	Proceedings of 1990 International Conference on Spoken Language Processing, 19.16, pp.825-828, Kobe, 1990.	50
Cheveigné 90-2a	The MapSignal Remote Speech Editor	Alain de Cheveigné	ATR Technical Report TR-I-0137	-
Cheveigné 90-2b	Experiments in Pitch Extraction	Alain de Cheveigné	ATR Technical Report TR-I-0138	-
Cheveigné 90-2c	Auditory Nerve Fiber Spike Generation Model	Alain de Cheveigné	ATR Technical Report TR-I-0139	-

List of Technical Publications from Speech Processing Department
(November 1989 through December 1990)
(Bold titles indicate papers written in English)

Ref.ID	Title(題名)	Authors(著者)	Journal(掲載誌)	Page
Dang 90-1a	シフト不変型決定論的ボルツマンマシンによる音声認識 [Shift-invariant Deterministic Boltzmann Machines for Phoneme Recognition]	Jean-Claude Dang 田村 震一 沢井 秀文 [S. Tamura H. Sawai]	電子情報通信学会技術研究報告 SP89-98, pp.1-6, 1990. [IECIE Technical Report, SP89-98, pp.1-6, 1990]	54
Dang 90-1b	シフト不変型決定論的ボルツマンマシンによる音声認識 [Shift-invariant deterministic Boltzmann Machines for Phoneme Recognition]	Jean-Claude Dang 田村 震一 沢井 秀文 [S. Tamura H. Sawai]	ATR Technical Report TR-I-0130	
Fukuza wa 90-9	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 [Speaker Adaptation Using Identity Mapping]	福沢 圭二 沢井 秀文 杉山 雅英 [K. Fukuzawa H. Sawai M. Sugiyama]	日本音響学会 平成2年度秋季研究 発表会講演論文集, 1-8-16, pp.31- 32, 1990. [Proc. of ASJ Fall Meeting, 1-8-16, pp.31-32, 1990]	60
Fujiwar a 90-7	ホルマントを用いる音声合成のためのルールインタプリタの開発 [Development of a Rule Interpreter for Speech Synthesis by Rule using the Formant Synthesizer]	藤原 宏之 野村 康雄 海木 延佳 鬼頭 樟梧 山下 洋一 溝口 理一郎 [H. Fujiwara Y. Nomura N. Kaiki A. Kito Y. Yamashita R. Mizoguchi]	電子情報通信学会技術研究報告, SP90-29, pp.9-16, 1990. [IECIE Technical Report, SP90-29, pp.9-16, 1990]	62
Gurgen 90-11	Line Spectrum Pair-Frequency-based Distance Measures for Speech Recognition	Fikret Gurgen S. Sagayama S. Furui	Proceedings of 1990 International Conference on Spoken Language Processing, 13.1, pp.521-524, Kobe, 1990	70
Hanaza wa 89-12	HMM-LR 音声認識システムの性能評価 [HMM-LR Speech Recognition System Performance]	花沢 利行 川端 泰 北 研二 中村 哲 鹿野 清宏 [T. Hanazawa T. Kawabata K. Kita S. Nakamura K. Shikano]	電子情報通信学会 技術研究報告, SP89-94, pp.63-70, 1989. [IECIE Technical Report, SP89-94, pp.63-70, 1989]	74
Hanaza wa 90-2a	Hidden Markov Model による音韻認識実験の結果 [Phoneme Recognition Using Hidden Markov Models]	花沢 利行 川端 泰 鹿野 清宏 [T. Hanazawa T. Kawabata K. Shikano]	ATR Technical Report TR-I-0147	

List of Technical Publications from Speech Processing Department
(November 1989 through December 1990)
(Bold titles indicate papers written in English)

Ref.ID	Title(題名)	Authors(著者)	Journal(掲載誌)	Page
Hanaza wa 90-2b	HMM音韻認識における音韻連鎖統計情報の利用 [HMM Phoneme Recognition Using Syllable Trigrams]	花沢 利行 川端 泰 伊藤 克巨 鹿野 清宏 [T. Hanazawa T. Kawabata K. Ito K. Shikano]	ATR Technical Report TR-I-0148	
Hanaza wa 90-3	HMM音韻認識における音節連鎖統計情報の利用 [HMM Phoneme Recognition Using Syllable Trigrams]	花沢 利行 川端 泰 伊藤 克巨 鹿野 清宏 [T. Hanazawa T. Kawabata K. Itoh K. Shikano]	日本音響学会 平成2年度 春季研究 発表会講演論文集, 3-3-9, pp.87- 88, 1990. [Proc. of ASJ Spring Meeting, 3-3- 9, pp.87-88, 1990]	82
Hanaza wa 90-4	ATR HMM-LR Continuous Speech Recognition System	T. Hanazawa K. Kita T. Kawabata S. Nakamura K. Shikano	Proceedings of 1990 International Conference on Acoustics, Speech, and Signal Processing, 52.4, pp.53- 56, Albuquerque, 1990	84
Hanaza wa 90-10	HMM-LR 音韻認識システムの性能評価 [HMM-LR Speech Recognition System Performance]	花沢 利行 北 研二 中村 哲 川端 泰 鹿野 清宏 [T. Hanazawa K. Kita S. Nakamura T. Kawabata K. Shikano]	日本音響学会誌, 46巻, 10号, pp.817-823, 1990. [The Journal of the Acoustical Society of Japan, Vol.46, 10, pp.817-823, 1990]	88
Hataza ki 90-1	スペクトログラムリーディング知識を用いた音韻セグメンテーションエキスパートシステム [Phoneme Segmentation Expert System Using Spectrogram Reading Knowledge]	畑崎 香一郎 小森 康弘 川端 泰 鹿野 清宏 [K. Hatazaki Y. Komori T. Kawabata K. Shikano]	電子情報通信学会論文誌, D-II, Vol.J73-D-II, No.1, pp.1-9, 1990. [Trans. of IEICE, D-II, Vol.J73-D-II, No.1, pp.1-9, 1990]	95
Hattori 89-12	話者重畳型HMMによる分節認識 [Speech Recognition Using Supplemented HMM]	服部 浩明 中村 哲 鹿野 清宏 [H. Hattori S. Nakamura K. Shikano]	電子情報通信学会 技術研究報告, SP89-90, pp.31-38, 1989. [IECIE Technical Report, SP89-90, pp.31-38, 1989]	104

List of Technical Publications from Speech Processing Department
(November 1989 through December 1990)
(Bold titles indicate papers written in English)

Ref.ID	Title(題名)	Authors(著者)	Journal(掲載誌)	Page
Hattori 90-3	話者適応における複数標準話者への重み付け [Speaker Weighted Supplemented HMM for Compensation of Speaker Articulatory Variations]	服部 浩明 中村 哲 鹿野 清宏 [H. Hattori S. Nakamura K. Shikano]	日本音響学会 平成2年度 春季研究発表会講演論文集, 2-3-1, pp.51-52, 1990. [Proc. of ASJ Spring Meeting, 2-3-1, pp.51-52, 1990]	112
Hattori 90-4	Supplementation of HMM for Articulatory Variation in Speaker Adaptation	H. Hattori S. Nakamura K. Shikano	Proceedings of 1990 International Conference on Acoustics, Speech, and Signal Processing, 53.6, pp.153-156, Albuquerque, 1990	114
Hattori 90-9	コード遷移確率に基づく学習データ重み付けによる話者適応化 [Speaker Adaptation with Training Sample Weighting Based on Markov Representation of Speaker]	服部 浩明 嵯峨山茂樹 [H. Hattori S. Sagayama]	日本音響学会 平成2年度 秋季研究発表会講演論文集, 1-8-15, pp.29-30, 1990. [Proc. of ASJ Fall Meeting, 1-8-15, pp.29-30, 1990]	118
Hattori 90-11	Speaker Weighted Training of HMM Using Multiple Reference Speakers	H. Hattori S. Nakamura K. Shikano S. Sagayama	Proceedings of 1990 International Conference on Spoken Language Processing, 5.6, pp.149-152, Kobe, 1990.	120
Hirato 90-3	種々の発話様式における韻律制御の検討 [Prosody Controls for Various Speaking Styles]	平戸 毅 匂坂 芳典 [T. Hirato Y. Sagisaka]	ATR Technical Report TR-I-0149	-
Honma 90-3	音素環境クラスタリングに基づいた音素単位HMMによる単語認識 [Word Recognition by Phoneme HMM Based on Phoneme Environment Clustering]	本間 茂 嵯峨山 茂樹 [S. Honma S. Sagayama]	日本音響学会 平成2年 春季研究発表会講演論文集, 1-3-13, pp.25-26, 1990. [Proc. of ASJ Fall Meeting, 1-3-13, pp.25-26, 1990]	124
Huber 90-11	Prosodic Transfer in Spoken Language Interpretation	Dieter Huber	Proceedings of 1990 International Conference on Spoken Language Processing, 12.7, pp.509-512, Kobe, 1990.	126
Inagaki 90-8	品詞情報による音韻継続時間長の分析 [Analysis of Segmental Duration Using Grammatical Information]	稲垣 英浩 匂坂 芳典 海木 延佳 [H. Inagaki Y. Sagisaka N. Kaiki]	ATR Technical Report TR-I-172	-
Iwai 90-3	F ₀ パターンと韻律構造上の「単位」について [A Study on F ₀ Patterns and Prosodic Units]	岩井 康雄 匂坂 芳典 [Y. Iwai Y. Sagisaka]	ATR Technical Report TR-I-160	-

List of Technical Publications from Speech Processing Department
(November 1989 through December 1990)
(Bold titles indicate papers written in English)

Ref.ID	Title(題名)	Authors(著者)	Journal(掲載誌)	Page
Kaiki 90-3	統計的手法を用いた文音声における音韻継続時間設定 [Phoneme Duration Setting in Sentence Utterances using Statistical Method]	海木 延佳 安部 勝雄 武田 一哉 匂坂 芳典 [N. Kaiki K. Abe K. Takeda Y. Sagisaka]	日本音響学会 平成2年度 春季研究発表会講演論文集, 1-4-13, pp.203-204, 1990. [Proc. of ASJ Spring Meeting, 1-4-13, pp.203-204, 1990]	130
Kaiki 90-5	文音声における音韻継続時間長の設定 [Phoneme Duration Setting in Sentence Utterances]	海木 延佳 武田 一哉 匂坂 芳典 [N. Kaiki K. Takeda Y. Sagisaka]	電子情報通信学会技術研究報告, SP90-2, pp.9-16, 1990. [IEICE Technical Report, SP90-2, pp.9-16, 1990]	132
Kaiki 90-9a	文音声における子音継続長の設定 [Consonant Duration Setting in Sentence Utterances]	海木 延佳 匂坂 芳典 [N. Kaiki Y. Sagisaka]	日本音響学会 平成2年度 秋季研究発表会講演論文集, 2-6-20, pp.259-260, 1990. [Proc. of ASJ Fall Meeting, 2-6-20, pp.259-260, 1990]	140
Kaiki 90-9b	The Control of Segmental Duration in Speech Synthesis Using Linguistic Properties	N. Kaiki K. Takeda Y. Sagisaka	Proceedings of the ESCA Workshop on Speech Synthesis, pp.165-168, Autrans, France, 1990.	142
Kaiki 90-11	Statistical Analysis for Segmental Duration Rules in Japanese Speech Synthesis	N. Kaiki K. Takeda Y. Sagisaka	Proceedings of 1990 International Conference on Spoken Language Processing, 1.5, pp.17-20, Kobe, 1990.	146
Kawabata 89-11a	構成的ニューラルネットワークによる音声認識 [Constructive Neural Network for Speech Recognition]	川端 素 [T. Kawabata]	ATR Technical Report TR-I-0122	-
Kawabata 89-11b	音韻モデルと文法を融合した音声認識 [Speech Recognition by Combination of Phoneme Models and Grammar]	川端 素 北 研二 [T. Kawabata K. Kita]	ATR ジャーナル, No.6, pp.6-9, 1989. [ATR Journal, No.6, pp.6-9, 1989]	150
Kawabata 90-1	HMM音韻認識における音節連鎖統計情報の利用 [HMM Phone Recognition Using Syllable Trigrams]	川端 素 花沢 利行 伊藤 克亘 鹿野 清宏 [T. Kawabata T. Hanazawa K. Itoh K. Shikano]	電子情報通信学会 技術研究報告, SP89-110, pp.7-12, 1989. [IEICE Technical Report, SP89-110, pp.7-12, 1989]	154

List of Technical Publications from Speech Processing Department
(November 1989 through December 1990)
(Bold titles indicate papers written in English)

Ref.ID	Title(題名)	Authors(著者)	Journal(掲載誌)	Page
Kawabata 90-3	k-近傍内挿学習による音韻認識 [Generalization Effects of k-Neighbor Interpolation Training]	川端 豪 [T. Kawabata]	日本音響学会平成2年度春季研究発表会講演論文集, 2-P-21, pp.161-162, 1990. [Proc. of ASJ Spring Meeting, 2-P-21, pp.161-162, 1990]	160
Kitagawa 90-1	HMM音韻認識における音韻連鎖統計情報の利用 [On the Use of Statistical Information of Phoneme Sequences in HMM-based Phoneme Recognition]	北川 英一郎 伊藤 克亘 川端 豪 鹿野 清宏 [E. Kitagawa K. Itoh T. Kawabata K. Shikano]	ATR Technical Report TR-I-0131	-
Komori 89-12	音韻認識エキスパートシステムにおける知識とTDNNの融合法 [Combining Time Delay Neural Networks (TDNN) and Spectrogram Reading Knowledge into a Phoneme Recognition Expert System]	小森 康弘 畑崎 香一郎 川端 豪 鹿野 清宏 [Y. Komori K. Hatazaki T. Kawabata K. Shikano]	電子情報通信学会技術研究報告, SP89-84, pp.63-70, 1989. [IEICE Technical Report, SP89-84, pp.63-70, 1989]	162
Komori 90-1	スペクトログラムリーディング知識とニューラルネットワークを用いた音韻認識エキスパートシステム [Phoneme Recognition Expert System Using Spectrogram Reading Knowledge and Neural Networks]	小森 康弘 畑崎 香一郎 川端 豪 鹿野 清宏 [Y. Komori K. Hatazaki T. Kawabata K. Shikano]	電子情報通信学会論文誌, D-II, Vol.J73-D-II, No.1, pp.10-18, 1990. [Trans. of IEICE, Vol.J73-D-II, No.1, pp.10-18, 1990]	170
Komori 90-3a	音韻認識エキスパートシステムにおける母音認識-TDNNによる母音スポッティング [Integrating Vowel-spotting TDNN into a Phoneme Recognition Expert System Using Spectrogram Reading Knowledge]	小森 康弘 畑崎 香一郎 川端 豪 鹿野 清宏 [Y. Komori K. Hatazaki T. Kawabata K. Shikano]	日本音響学会平成2年度春季研究発表会講演論文集, 2-P-18, pp.155-156, 1990. [Proc. of ASJ Spring Meeting, 2-P-18, pp.155-156, 1990]	179
Komori 90-3b	時間構造を考慮したニューラルネットワークによる音韻認識 [Phoneme Identification Neural Networks Concerning Phonetic Temporal Structure]	小森 康弘 南 泰浩 鹿野 清宏 [Y. Komori Y. Minami K. Shikano]	日本音響学会平成2年度春季研究発表会講演論文集, 2-P-19, pp.157-158, 1990. [Proc. of ASJ Spring Meeting, 2-P-19, pp.157-158, 1990]	181
Komori 90-4	Combining Phoneme Identification Neural Networks into an Expert System Using Spectrogram Reading Knowledge	Y. Komori K. Hatazaki T. Tanaka T. Kawabata	Proceedings of 1990 International Conference on Acoustics, Speech, and Signal Processing, S10.5, pp.505-508, Albuquerque, 1990.	183

List of Technical Publications from Speech Processing Department
(November 1989 through December 1990)
(Bold titles indicate papers written in English)

Ref.ID	Title(題名)	Authors(著者)	Journal(掲載誌)	Page
Komori 90-11	A Fuzzy Training Approach for Phoneme Classification Neural Networks	Y. Komori S. Sagayama A. Waibel	ATR Technical Report TR-I-190	-
Kurematsu 89-10	ATRにおける自動翻訳の概要 [Overview of ATR Researches into Telephone Interpretation]	樽松 明 鹿野 清宏 川端 豪 飯田 仁 森元 暉 [A. Kurematsu K. Shikano T. Kawabata H. Iida T. Morimoto]	ATR Technical Report TR-I-184	-
Kurematsu 90-1	自動翻訳電話の可能性 [On the Feasibility of Automatic Interpreting Telephony]	樽松 明 [A. Kurematsu]	日本ビジネスレポート 技術予測シリーズ pp.165-174, 1990. [Technical Forecast Series, Japan Business Report, pp.165-174, 1989]	187
Kurematsu 90-3	ATR Japanese Speech Database as a Tool of Speech Recognition and Synthesis	A. Kurematsu K. Takeda Y. Sagisaka S. Katagishi H. Kuwabara K. Shikano	Speech Communication, Vol.9, pp.357-363, 1990, North-Holland.	197
Kurematsu 90-5	ATRにおける音声認識の研究 [Research on Speech Recognition at ATR]	樽松 明 [A. Kurematsu]	東北大学シンポジウム 音声の自動認識の現状と将来 pp.1-7, 1990. [Proceedings of Tohoku University Symposium on Automatic Speech Recognition Present and Future, pp.1-7, 1990]	204
Kurematsu 90-9	音声言語の理解における概念形成の課題 [Problems on Concept Formulation in Spoken Language Understanding]	樽松 明 [A. Kurematsu]	電子情報通信学会第二種研究会, 「言語獲得-概念形成」 LA90-6, pp.1-10, 1990.	211
Kurematsu 90-10	自動翻訳電話--異言語間のコミュニケーションを目指して [Automatic Telephone Interpretation -Towards a Communication Between Different Languages]	樽松 明 [A. Kurematsu]	電子情報通信学会技術研究報告, HC90-15, pp.1-8, 1990. [IEICE Technical Report, HC90-15, pp.1-8, 1990]	221
Kurematsu 90-11	A Perspective of Telephone Interpretation Research	A. Kurematsu	Proceedings of Pacific Rim International Conference on Artificial Intelligence, pp.11-16, 1990.	229

List of Technical Publications from Speech Processing Department
(November 1989 through December 1990)
(Bold titles indicate papers written in English)

Ref.ID	Title(題名)	Authors(著者)	Journal(掲載誌)	Page
Kuwabara 90-8	Voice Quality Control through Vector Quantization	H. Kuwabara M. Abe	ISSPA(International Symposium on Signal Processing and Its Applications), 1990	235
Maruyama 89-11	HMM音韻連結学習とNET gramを用いた英単語音声の認識 [English Word Recognition using HMM Phone Concatenated Training and Netgram]	丸山 活輝 中村 雅己 川端 豪 鹿野 清宏 [K. Maruyama M. Nakamura T. Kawabata K. Shikano]	ATR Technical Report TR-I-0123	-
Maruyama 89-	HMM Based Word Recognition using Word Category Prediction Neural Network	K. Maruyama M. Nakamura T. Kawabata K. Shikano	The Journal of the Acoustical Society of America, 568, CC6, Supplement 1, Vol.86, Fall 1989	239
Maruyama 89-12	HMM音韻連結学習とNET gramを用いた英単語音声の認識 [English Word Recognition using HMM Phone Concatenated Training and Netgram]	丸山 活輝 中村 雅己 川端 豪 鹿野 清宏 [K. Maruyama M. Nakamura T. Kawabata K. Shikano]	電子情報通信学会 技術研究報告 SP89-89, pp.25-30, 1989 [IECIE Technical Report, SP89-89, pp.25-30, 1989]	240
Maruyama 90-3a	NET gramを用いたHMM英単語音声認識の改善 [Improvement of HMM-based English Word Recognition Using NETgram]	丸山 活輝 中村 雅己 川端 豪 鹿野 清宏 [K. Maruyama M. Nakamura T. Kawabata K. Shikano]	日本音響学会 平成2年度 春季研究発表会講演論文集, 3-3-7, pp.83-84, 1990. [Proc. of ASJ Spring Meeting, 3-3-7, pp.83-84, 1990]	246
Maruyama 90-3b	NET gramを用いたHMM英単語音声認識の改善 [Improvement of HMM-based English Word Recognition Using NETgram]	丸山 活輝 中村 雅己 川端 豪 鹿野 清宏 [K. Maruyama M. Nakamura T. Kawabata K. Shikano]	ATR Technical Report TR-I-0133	-
Matsunaga 90-11a	Sentence Speech Recognition using Semantic Dependency Analysis	S. Matsunaga S. Sagayama	Proceedings of 1990 International Conference on Spoken Language Processing, 21.9, pp.929-932, Kobe, 1990.	248
Matsunaga 90-11b	A Continuous Speech Recognition System Based on a Two-level Grammar Approach	S. Matsunaga S. Sagayama S. Honma S. Furui	Proceedings of 1990 International Conference on Spoken Language Processing, S11.7, pp.589-592, Kobe, 1990.	252

List of Technical Publications from Speech Processing Department
(November 1989 through December 1990)
(Bold titles indicate papers written in English)

Ref.ID	Title(題名)	Authors(著者)	Journal(掲載誌)	Page
Minami 90-1	TDNN音韻スポッティングと予測LRパーザを用いた大語彙単語音声認識 [Large Vocabulary Spoken Word Recognition Using Time-Delay Neural Network Phoneme Spotting and Predictive LR-Parsing]	南 泰浩 沢井 秀文 宮武 正典 [Y. Minami H. Sawai M. Miyatake]	電子情報通信学会 技術研究報告 SP89-99, pp.7-13, 1990. [IECIE Technical Report, SP89-99, pp.7-13, 1990]	256
Minami 90-2a	TDNN音韻スポッティングと予測LRパーザを用いた大語彙単語音声認識 [Large Vocabulary Spoken Word Recognition Using Time-Delay Neural Network Phoneme Spotting and Predictive LR-Parsing]	南 泰浩 沢井 秀文 宮武 正典 鹿野 清宏 [Y. Minami H. Sawai M. Miyatake K. Shikano]	ATR Technical Report TR-I-0144	-
Minami 90-2b	TDNNの構造の音韻認識率、シフトインバリエント性への影響 [The Effect of TDNN Structures on Phoneme Recognition Rates and Shift Invariance]	南 泰浩 沢井 秀文 [Y. Minami H. Sawai]	ATR Technical Report TR-I-0145	-
Minami 90-3	入力層・中間層におけるベクトルの近傍の情報を利用したTDNN出力の平滑化 [Output Smoothing for TDNN Using Information on Input Vectors Neighborhood Input Layer]	南 泰浩 田村 慶一 沢井 秀文 鹿野 清宏 [Y. Minami S. Tamura H. Sawai K. Shikano]	日本音響学会 平成2年度 春季研究発表会講演論文集, 1-3-18, pp.35-36, 1990. [Proc. of ASJ Spring Meeting, 1-3-18, pp.35-36, 1990]	263
Minami 90-6	時間遅れ神経回路網(TDNN)による音韻スポッティング法と予測LRパーザを用いた大語彙単語音声認識 [Large Vocabulary Spoken Word Recognition Using Time-Delay Neural Network Phoneme Spotting and Predictive LR-Parsing]	南 泰浩 沢井 秀文 宮武 正典 [Y. Minami H. Sawai M. Miyatake]	電子情報通信学会論文誌D-II, Vol.J73-D-II, No.6, PP.788-795, 1990. [Trans. of IEICE, Vol.J73-D-II, No.6, PP.788-795, 1990]	265
Minami 90-10	発話変動にロバストなTDNNの検討 [A Study on Robust TDNNs against the Variability of Speaking Styles]	南 泰浩 沢井 秀文 [Y. Minami H. Sawai]	ATR Technical Report TR-I-183	-
Mimura 90-9	統計的手法を用いた文音声における振幅制御 [Power Control in Sentence Utterances Using Statistical Method]	三村 克彦 海木 延佳 匂坂 芳典 [K. Mimura Y. Kaiki Y. Sagisaka]	日本音響学会 平成2年度秋季研究発表会講演論文集, 2-6-19, pp.257-258, 1990. [Proc. of ASJ Fall Meeting, 2-6-19, pp.257-258, 1990]	273

List of Technical Publications from Speech Processing Department
(November 1989 through December 1990)
(Bold titles indicate papers written in English)

Ref.ID	Title(題名)	Authors(著者)	Journal(掲載誌)	Page
Miyatake 90-4	Integrated Training for Spotting Japanese Phonemes Using Large Phonemic Time-Delay Neural Networks	M. Miyatake H. Sawai Y. Minami K. Shikano	Proceedings of 1990 International Conference on Acoustics, Speech, and Signal Processing, 58.10, pp.449-452, Albuquerque, 1990.	275
Miyatake 90-5	時間遅れ神経回路網(TDNN)による音韻スポッティングのための学習法とその効果 [Training Methods and Their Effects for Spotting Japanese Phonemes Using Time-Delay Neural Networks]	宮武 正典 沢井 秀文 鹿野 清宏 [M. Miyatake H. Sawai K. Shikano]	電子情報通信学会論文誌, D-II, Vol.J73-D-II, No.5 pp.699-706, 1990. [Trans. of IEICE, Vol.J73-D-II, No.5, pp.699-706, 1990]	279
Miyatake 90-12	種々の発話様式に見られる韻律特徴とその制御 [Prosodic Characteristics and Their Control in Japanese Speech with Various Speaking Styles]	宮武 正典 匂坂 芳典 [M. Miyatake Y. Sagisaka]	電子情報通信学会論文誌, D-II, Vol.J73-D-II No.12, pp.1929-1935, 1990. [Trans. of IEICE, Vol.J73-D-II, No.12, pp.1929-1935, 1990]	287
Nagai 90-9	HMM-LR 法における音素文脈依存型LRパーザの検討 [Study on Phonemic Context Dependent LR Parser in HMM-LR Method]	永井 明人 北 研二 巖山 茂樹 [A. Nagai K. Kita S. Sagayama]	日本音響学会平成2年度秋季研究発表会講演論文集, 3-8-16, pp.125-126, 1990. [Proc. of ASJ Fall Meeting, 3-8-16, pp.125-126, 1990]	294
Nakamura M 89-10	ニューラルネットにおけるバックプロパゲーション学習の効率化方法 [A New Method to Speed up the Back-Propagation Algorithm]	中村 雅己 鹿野 清宏 [M. Nakamura K. Shikano]	ATR Technical Report TR-I-0119	-
Nakamura M 90-3	ニューラルネットによる音韻フィルター [A Study of Phoneme Filter Using Neural Network]	中村 雅己 田村 震一 [M. Nakamura S. Tamura]	日本音響学会平成2年度春季研究発表会講演論文集, 2-P-24, pp.167-168, 1990. [Proc. of ASJ Spring Meeting, 2-P-24, pp.167-168, 1990]	296
Nakamura M 90-6	ニューラルネットによる音素フィルタを用いた母音認識 [Vowel Recognition by Phoneme Filter Neural Networks]	中村 雅己 田村 震一 [M. Nakamura S. Tamura]	電子情報通信学会技術研究報告, SP90-11, pp.17-23, 1990 [IEICE Technical Report, SP90-11, 1990]	298
Nakamura M 90-8	Neural Network Approach to Word Category Prediction for English Texts	M. Nakamura K. Maruyama T. Kawabata K. Shikano	COLING'90, Computational Linguistics, Vol.3, pp.213-218, Finland, 1990.	305
Nakamura M 90-9a	ニューラルネットによる英単語品詞列予測モデル [English Word Category Prediction -Based on Neural Network]	中村 雅己 鹿野 清宏 [M. Nakamura K. Shikano]	ATR Technical Report TR-I-176	-

List of Technical Publications from Speech Processing Department
(November 1989 through December 1990)
(Bold titles indicate papers written in English)

Ref.ID	Title(題名)	Authors(著者)	Journal(掲載誌)	Page
Nakamura M 90-9b	ニューラルネットによる音素フィルタを用いた母音認識 [Vowel Recognition by Phoneme Filter Neural Networks]	中村 雅己 田村 震一 [M. Nakamura S. Tamura]	ATR Technical Report TR-I-177	-
Nakamura M 90-11	Vowel Recognition by Phoneme Filter Neural Networks	M. Nakamura S. Tamura	Proceedings of 1990 International Conference on Spoken Language Processing 16.3, pp.669-672, Kobe, 1990.	311
Nakamura M 90-9	時間遅れ神経回路網を用いた不特定話者の音韻認識 [Speaker-independent Phoneme Recognition Using Time-Delay Neural Networks]	中村 悟 沢井 秀文 [S. Nakamura H. Sawai]	ATR Technical Report TR-I-178	-
Nakamura M 90-12	不特定話者音素認識のためのニューラルネットアーキテクチャの検討 [A Preliminary Study on Neural Network Architectures for Speaker-Independent Phoneme Recognition]	中村 悟 沢井 秀文 [S. Nakamura H. Sawai]	電子情報通信学会技術研究報告 SP.90-61, pp.33-40, 1990. [IEICE Technical Report SP.90-61, pp.33-40, 1990]	315
Nakamura S 90-3	ベクトル量子化話者適応化のTDNN音韻認識への適用 [VQ-based Speaker Adaptation Applied to TDNN Phoneme Recognition]	中村 哲 鹿野 清宏 [S. Nakamura K. Shikano]	日本音響学会平成2年度春季研究発表会講演論文集, 2-P-22, pp.163-164, 1990. [Proc. of ASJ Spring Meeting, 2-P-22, pp.163-164, 1990]	323
Nakamura S 90-4	A Comparative Study of Spectral Mapping for Speaker Adaptation	S. Nakamura K. Shikano	Proceedings of 1990 International Conference on Acoustics, Speech, and Signal Processing, 53.7, pp.157-160, Albuquerque, 1990.	325
Nakamura S 90-12	話者重畳型HMMを用いた日本語音韻認識における話者適応化の改善 [Improved Speaker Adaptation Using Speaker Supplemented HMM for Japanese Phoneme Recognition]	中村 哲 服部 浩明 鹿野 清宏 [S. Nakamura H. Hattori K. Shikano]	電子情報通信学会論文誌, D-II, Vol.J73-D-II, No.12, pp.1919-1928, 1990. [Trans. of the IEICE, D-II, Vol.J73-D-II, No.12, pp.1919-1928, 1990]	329
Ohkura 90-6	コードブックマッピングを用いた雑音環境下での音声認識 [Speech Recognition of Under Noisy Environment by Codebook Mapping]	大倉 計美 服部 浩明 杉山 雅英 鹿野 清宏 [K. Ohkura H. Hattori M. Sugiyama K. Shikano]	電子情報通信学会技術研究報告 SP90-12, pp.25-32, 1990. [IEICE Technical Report, SP90-12, pp.25-32, 1990.]	339

List of Technical Publications from Speech Processing Department
(November 1989 through December 1990)
(Bold titles indicate papers written in English)

Ref.ID	Title(題名)	Authors(著者)	Journal(掲載誌)	Page
Ohkura 90-9	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 [The Application of Noise Reduction Neural Network to Speech Recognition]	大倉 計美 杉山 雅英 [K. Ohkura M. Sugiyama]	日本音響学会平成2年度秋季研究 発表会講演論文集, 1-8-3, pp.5-6, 1990. [Proc. of ASJ Fall Meeting, 1-8-3, pp.5-6, 1990]	347
Rainton 90-9	Speech Analysis / Synthesis using a Partial outer Product Expansion of the Wigner distribution	David Rainton	日本音響学会平成2年度秋季研究 発表会講演論文集, 1-8-1, pp.1-2, 1990. [Proc. of ASJ Fall Meeting, 1-8-1, pp.1-2, 1990]	349
Rainton 90-10	Speech Analysis and Enhancement using the Time-frequency Wigner distribution	David Rainton	電子情報通信学会技術研究報告, SP90-49, pp.39-46, 1990. [IECIE Technical Report, SP90-49, pp.39-46, 1990]	351
Rainton 90-11	Time-Frequency Spectral Analysis of Speech	David Rainton S.J. Young	Proceedings of 1990 International Conference on Spoken Language Processing, 9.1, pp.349-352, 1990.	359
Sadano bu 90-4	合成的字音語のアクセントの 字数 [Accent of Sino-Japanese Complex Word and the Number of Charcters]	定延 利之 匂坂 芳典 [T. Sadanobu Y. Sagisaka]	ATR Technical Report TR-I-158	-
Sagaya ma 90-4	音素環境クラスティングの HMM大語彙音声認識への適 用 [Phoneme Environment Clustering Applied to Large Vocabulary Speech Recognition]	嵯峨山 茂樹 [S. Sagayama]	文部省科研費総合研究(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5回研究討 論会 榊原温泉 H2.4.20 [5th meeting of the committee of New Approaches for Combining Markov Models and Neural Networks, Sakakibara-onsen, April, 1990]	363
Sagaya ma 90-7	音声認識におけるいくつか の基本問題について [Basic Problems in Speech Recognition]	嵯峨山 茂樹 [S. Sagayama]	文部省科研費総合研究(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! シンポジウム (湯河原) H2.7.20 [Symposium on New Approaches for Combining Markov Models and Neural Networks, Yugawara- onsen, July, 1990]	367
Sagaya ma 90-11	Estimation of Unknown Context Using a Phoneme Environment Clustering Algorithm	S. Sagayama S. Honma	Proceedings of 1990 International Conference on Spoken Language Processing, 9.4, pp.361-364, 1990.	371

List of Technical Publications from Speech Processing Department
(November 1989 through December 1990)
(Bold titles indicate papers written in English)

Ref.ID	Title(題名)	Authors(著者)	Journal(掲載誌)	Page
Sagaya ma 90-12	行列演算によるHMM音声認 識アルゴリズムの表現につ いて [A Matrix Representation of HMM-Based Speech Recognition Algorithms]	嵯峨山 茂樹 [S. Sagayama]	電子情報通信学会技術研究報告, SP90-65, pp.63-70, 1990. [IECIE Technical Report, SP90-65, pp.63-70, 1990]	375
Sagisak a 89-11a	On the Unit Set Design for Speech Synthesis by Rule using Non-uniform Units	Y. Sagisaka	The Journal of the Acoustical Society of America, Supplement, Vol.86, 579, FF24, Fall, 1989.	383
Sagisak a 89-11b	「音声研究」日本音響学会座 談会 [Meeting on "Speech Research"]	匂坂 芳典, 他 [Y. Sagisaka]	日本音響学会 学会誌, 45巻, 11月 号, pp.861-871, 1989. [The Journal of the Acoustical Society of Japan, Vol.45, 11, pp.861-871, 1989]	384
Sagisak a 90-1	F ₀ ボタン概形制御の定量的検 討 [On the Quantification of Global F ₀ Pattern Control]	匂坂 芳典 [Y. Sagisaka]	電子情報通信学会 技術研究報告, SP89-111, pp.13-18, 1990. [IECIE Technical Report, SP89-111, pp.13-18, 1990]	395
Sagisak a 90-3	Speech Synthesis from Text	Y. Sagisaka	IEEE Communication Magazine, Special Issue, pp.35-41, 1990.	401
Sagisak a 90-4a	On the Prediction of Global F₀ Shape for Japanese Text- to-Speech	Y. Sagisaka	Proceedings of 1990 International Conference on Acoustics, Speech, and Signal Processing, 56a.9, pp.325-328, Albuquerque, 1990.	409
Sagisak a 90-4b	音声合成の立場からみた音声 処理単位 [Segment Unit for Speech Synthesis]	匂坂 芳典 [Y. Sagisaka]	文部省科研費総合研究(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5回研究討 論会 榊原温泉 H2.4.20 [Symposium on New Approaches for Combining Markov Models and Neural Networks, Sakakibara- onsen, April, 1990]	413
Sagisak a 90-	Phonotactic Constraintsを用 いた音声単位の構成法 [The design of Synthesis unit set Using Phonotactic Constraints]	匂坂 芳典 [Y. Sagisaka]	近畿音声言語研究会資料 [KINKI Society for Phonetics]	415
Sagisak a 90-9a	統語構造に基づくF ₀ 上昇現象 の統計的分析 [Statistic Analysis on F ₀ Boosting in Relation to Syntactic Structure]	匂坂 芳典 [Y. Sagisaka]	日本音響学会平成2年度秋季研究 発表会講演論文集, 2-6-15, pp.249- 250, 1990. [Proc. of ASJ Fall Meeting, 2-6-15, pp.249-250, 1990]	422

List of Technical Publications from Speech Processing Department
(November 1989 through December 1990)
(Bold titles indicate papers written in English)

Ref.ID	Title(題名)	Authors(著者)	Journal(掲載誌)	Page
Sagisaka 90-9b	Units and Rules for Speech Synthesis	Y. Sagisaka	Proceedings of the Tutorial Day Workshop on Speech Synthesis, pp.1-10, AuTrans, France, 1990.	424
Sagisaka 90-11	A Large-scale Japanese Speech Database	Y. Sagisaka K. Takeda M. Abe S. Katagishi T. Umeda H. Kuwabara	Proceedings of 1990 International Conference on Spoken Language Processing, 24.4, pp.1089-1092, Kobe, 1990.	433
Sawai 90-9a	TDNN-LR文節音声認識システムにおける追加学習の効果 [Effects of Incremental Training in the TDNN-LR Phrase Speech Recognition System]	沢井 秀文 [H. Sawai]	日本音響学会平成2年度秋季研究発表会講演論文集, 2-P-11, pp.151-152, 1990. [Proc. of ASJ Fall Meeting, 2-P-11, pp.151-152, 1990]	437
Sawai 90-9b	時間・周波数変動に強い時間遅れ神経回路網(TDNN) [Time-Frequency Shift-Tolerant Time-Delay Neural Networks]	沢井 秀文 [H. Sawai]	日本音響学会平成2年度秋季研究発表会講演論文集, 2-P-12, pp.153-154, 1990. [Proc. of ASJ Fall Meeting, 2-P-12, pp.153-154, 1990]	439
Sawai 90-10	時間遅れ神経回路網(TDNN)による音声認識 [Speech Recognition Using Time-Delay Neural Networks]	沢井 秀文 [H. Sawai]	電気関係学会関西支部連合大会予稿集, S14-2 [Proceedings of United Meeting of Kansai District in Electronics-related Societies, S14-2]	441
Sawai 90-11	The TDNN-LR Large-vocabulary and Continuous Speech Recognition System	H. Sawai	Proceedings of 1990 International Conference on Spoken Language Processing, 31.4, pp.1349-1352, Kobe, 1990.	442
Shikano 89-10	Research Activities of the Speech Processing Department	鹿野 清宏 [K. Shikano]	ATR Technical Report TR-I-0115	-
Shikano 89-12	ニューラルネットワークによる音声認識 [Speech Recognition Using Neural Networks]	鹿野 清宏 [K. Shikano]	コロナ社 コンピューートルール, No.29, pp.42-51, 1989. [Computrol, No. 29, pp.42-51, 1989]	446
Shikano 90-5	Approaches to Continuous Speech Recognition Using Time-Delay Neural Networks and Learning Vector Quantization	K. Shikano	日独情報技術フォーラム通産省、西ドイツ, 1990.5 [Deutsch-Japanisches Forum Informationstechnologie, Mai 1990]	456
Shimodaira 90-12	ビッチボタン連続整合による連続音声のセグメンテーション [Phrase Segmentation of Continuous Speech]	下平 博 嵯峨山 茂樹 木村 正行 [H. Shimodaira S. Sagayama M. Kimura]	電子情報通信学会技術研究報告, SP90-72, pp.33-40, 1990. [IECIE Technical Report, SP90-72, pp.33-40, 1990]	461

List of Technical Publications from Speech Processing Department
(November 1989 through December 1990)
(Bold titles indicate papers written in English)

Ref.ID	Title(題名)	Authors(著者)	Journal(掲載誌)	Page
Sugiyama 90-5	歪み尺度測地線を用いた音声スペクトルの補間 [Spectral Interpolation using distortion Geodesic Lines]	杉山 雅英 [M. Sugiyama]	電子情報通信学会技術研究報告, SP90-3, pp.17-24, 1990. [IECIE Technical Report, SP-90-3, pp.17-24, 1990]	469
Sugiyama 90-8	Automatic Language Recognition using Acoustic Features	M. Sugiyama	ATR Technical Report TR-I-167	-
Sugiyama 90-9	ニューラルネットによる集合間写像の教師なし学習 [Unsupervised Training Methods for Set Mappings Using Neural Networks]	杉山 雅英 福沢 圭二 沢井 秀文 嵯峨山 茂樹 [M. Sugiyama K. Fukuzawa H. Sawai S. Sagayama]	日本音響学会平成2年度秋季研究発表会講演論文集, 2-P-10, pp.149-150, 1990. [Proc. of ASJ Fall Meeting, 2-P-10, pp.149-150, 1990]	477
Sugiyama 90-10a	ATRにおけるNeural Networkを用いた音声情報処理 [Neural Networks Applied to Speech Processing in ATR]	杉山 雅英 [M. Sugiyama]	ATR Technical Report TR-I-173	-
Sugiyama 90-10b	ニューラルネットワークを用いた音声情報処理 [Speech Information Processing Using Neural Networks]	杉山 雅英 [M. Sugiyama]	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」pp.1-16, 1990. [Draft for Seminar of Neural Network and Information Science, New Trends of Neural Network Models, pp.1-16, 1990]	479
Sugiyama 90-11	Spectral Interpolation Using Distortion Geodesic Lines	M. Sugiyama	Proceedings of 1990 International Conference on Spoken Language Processing, 11.15, pp.477-480, Kobe, 1990.	503
Takahashi 90-11	Isolated Word Recognition Using Pitch Pattern Information	S. Takahashi S. Matsunaga S. Sagayama	Proceedings of 1990 International Conference on Spoken Language Processing, 13.9, pp.553-556, Kobe, 1990.	507
Takeda 90-1	エキスパートシステムを用いた単位選択の検討 [A Rule-based Approach for Synthesis Unit Selection]	武田 一哉 匂坂 芳典 安部 勝雄 [K. Takeda Y. Sagisaka M. Abe]	電子情報通信学会技術研究報告 SP89-113, pp.27-32, 1990. [IECIE Technical Report, SP89-113, pp.27-32, 1990]	511

List of Technical Publications from Speech Processing Department
(November 1989 through December 1990)
(Bold titles indicate papers written in English)

Ref.ID	Title(題名)	Authors(著者)	Journal(掲載誌)	Page
Takeda 90-2a	複合音声単位を用いる規則合成実験システム [Speech Synthesis System Using Non-Uniform Units]	武田 一哉 安部 勝雄 匂坂 芳典 海木 延佳 [K. Takeda M. Abe Y. Sagisaka N. Kaiki]	ATR Technical Report TR-I-0140	-
Takeda 90-2b	種々の音韻連接単位を用いる規則合成方式の診断的な評価 [Evaluation and Diagnosis of Selective Use of Non-uniform Units for Speech Synthesis-by-rule]	武田 一哉 阿部 匡伸 匂坂 芳典 [K. Takeda K. Abe Y. Sagisaka]	ATR Technical Report TR-I-0142	-
Takeda 90-2c	大規模音声データベースに基づく音声合成 [Automatic Rule Derivation for Speech Synthesis from Large-scaled Speech Database]	武田 一哉 [K. Takeda]	ATR Technical Report TR-I-0143	-
Takeda 90-3	合成単位の音響的屬性と合成音との関係 [On the Acoustic-phonetic Properties of Synthesis Units and the Resulting Synthesized Speech]	武田 一哉 安部 勝雄 匂坂 芳典 [K. Takeda K. Abe Y. Sagisaka]	日本音響学会平成2年度春季研究発表会講演論文集, 1-4-14, pp.205-206, 1990. [Proc. of ASJ Spring Meeting, 1-4-14, pp.205-206, 1990]	517
Takeda 90-9	On Unit Selection Algorithms and Their Evaluation in Non-uniform Speech Synthesis	K. Takeda K. Abe Y. Sagisaka	Proceedings of the Tutorial and Research Workshop on Speech Synthesis, pp.35-38, France, 1990.	519
Takeda 90-11	On the Unit Search Criteria and Algorithms for Speech Synthesis using Non-uniform Units	K. Takeda K. Abe Y. Sagisaka	Proceedings of 1990 International Conference on Spoken Language Processing, 8.8, pp.341-344, Kobe, 1990.	523
Takeda 90-12	選択的に合成単位を用いる規則音声合成 [Speech Synthesis by Rule Based on Adaptive Unit Selection]	武田 一哉 安部 勝雄 匂坂 芳典 [K. Takeda K. Abe Y. Sagisaka]	電子情報通信学会論文誌, Vol. J73-D-II, No.12 pp.1945-1951, 1990. [Trans. of the IEICE, Vol. J73-D-II, No.12 pp.1945-1951, 1990]	527
Takami 90-6	対判定型TDNNによる音素認識 [Phoneme Recognition by Pairwise Discriminant TDNN]	鷹見 淳一 嵯峨山 茂樹 [J. Takami S. Sagayama]	電子情報通信学会技術研究報告 SP90-10, pp.9-16, 1990. [IECIE Technical Report, SP90-10, pp.9-16, 1990]	534

List of Technical Publications from Speech Processing Department
(November 1989 through December 1990)
(Bold titles indicate papers written in English)

Ref.ID	Title(題名)	Authors(著者)	Journal(掲載誌)	Page
Takami 90-9	対判定型TDNNにおける中間値学習の効果 [Effects of the Training with Middle Value on Pairwise Discriminant TDNN]	鷹見 淳一 嵯峨山 茂樹 [J. Takami S. Sagayama]	日本音響学会平成2年秋季研究発表会講演論文集, 2-P-15, pp.159-160, 1990. [Proc. of ASJ Fall Meeting, 1990]	542
Takami 90-11	Phoneme Recognition by Pairwise Discriminant TDNNs	J. Takami S. Sagayama	Proceedings of 1990 International Conference on Spoken Language Processing, 16.5, pp.677-680, Kobe, 1990.	544
Tamura 89-10	フィードフォワードニューラルネットワークの解釈 [On Interpretations of a Feedforward Neural Network]	田村 震一 [S. Tamura]	ATR Technical Report TR-I-0116	-
Tamura 90-2	波形入出力による雑音抑圧ニューラルネットワークの最適化 [Improvements to the Noise Reduction Neural Network]	田村 震一 中村 雅己 [S. Tamura M. Nakamura]	電子情報通信学会技術研究報告, SP89-120, pp.9-14, 1990. [IECIE Technical Report SP89-120, pp.9-14, 1990]	548
Tamura 90-3	波形入出力による雑音抑圧ニューラルネットワークの改良 [Improvements to the Noise Reduction Neural Network]	田村 震一 中村 雅己 [S. Tamura M. Nakamura]	日本音響学会平成2年度春季研究発表会講演論文集, 3-4-18, pp.303-304, 1990. [Proc. of ASJ Spring Meeting, 3-4-18, pp.303-304, 1990]	554
Tamura 90-4	Improvements to the Noise Reduction Neural Network	S. Tamura M. Nakamura	Proceedings of 1990 International Conference on Acoustics, Speech, and Signal Processing, S15b.5, pp.825-828, Albuquerque, 1990.	556
Umeda 90-8	声質変換技術と高品質ピッチ変換法 [Voice Conversion Techniques and a High-quality Pitch Conversion Algorithm]	梅田 哲夫 [T. Umeda]	ATR Technical Report TR-I-175	-
Yamashita 90-11	A Support Environment Based on Rule Interpreter for Synthesis by Rule	Y. Yamashita H. Fujiwara Y. Nomura N. Kaiki R. Mizoguchi	Proceedings of 1990 International Conference on Spoken Language Processing, 19.2, pp.769-772, Kobe, 1990.	560
Yoshida 90-5	母音無声化の要因分析 [Factor Analysis of Vowel devoicing in Japanese]	吉田 夏也 匂坂 芳典 [N. Yoshida Y. Sagisaka]	ATR Technical Report TR-I-159	-
Waibel 89-11	Connectionist Large Vocabulary Word Recognition	Alex Waibel	ATR Technical Report TR-I-0120	-

97

List of Technical Publications from Speech Processing Department

(November 1989 through December 1990)

(Bold titles indicate papers written in English)

Ref.ID	Title(題名)	Authors(著者)	Journal(掲載誌)	Page
Waibel 89-12	Modularity and Scaling in Large Phonemic Neural Network	Alex Waibel H. Sawai K. Shikano	IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol.37, No.12, pp.1888- 1898, 1989.	564

List of Publications, January through December, 1991

[Abe91SP2] 1
 阿部 匡伸, 嵯峨山 茂樹: “音素セグメントを単位とする声質変換,” 電子情報通信学会技術研究報告, SP90-88, pp. 25-32 (1991.02).
 (Masanobu Abe, Shigeki Sagayama: “A Phoneme Segment Based Approach to Voice Conversion,” *IEICE Technical Report, SP90-88, pp. 25-32 (1991.02).*)

[Abe91ASJ3] 9
 阿部 匡伸, 嵯峨山 茂樹: “音素セグメントを単位とする声質変換の評価,” 日本音響学会平成3年度春季研究発表会講演論文集, 2-6-10, pp. 265-266 (1991.03).
 (Masanobu Abe, Shigeki Sagayama: “An Evaluation of Phoneme Segment Approach to Voice Conversion,” *ASJ Spring Meeting, 2-6-10, pp. 265-266 (1991.03).*)

[Abe91ICASSP5] 11
 Masanobu Abe: “A Segment-based Approach to Voice Conversion,” Proc. of 1991 International Conference on Acoustics, Speech, and Signal Processing, S11.8, pp. 765-768, Toronto, Canada (1991.05).

[Abe91JASA7] 15
 Masanobu Abe, Kiyohiro Shikano, Hisao Kuwabara: “Statistical Analysis of Bilingual Speaker’s Speech for Cross-language Voice Conversion,” *The Journal of the Acoustical Society of America*, Vol. 90, No. 1, pp. 76-82 (1991.07).

◆

[Campbell91SP1] 22
 W. N. Campbell: “英語音声合成システムのための階層的なタイミング制御,” 電子情報通信学会技術研究報告, SP90-77, pp. 1-6 (1991.01).
 (W. N. Campbell: “Higher-Level Timing Control for an English-Language Speech Synthesis System,” *IEICE Technical Report, SP90-77, pp. 1-6 (1991.01).*)

[Campbell91JPhon1] 28
 W. N. Campbell, S. D. Isard: “Segment Durations in a Syllable Frame,” *Journal of Phonetics*, Vol. 19, No. 1, Special Issue on Speech Synthesis, pp. 37-47 (1991.01).

[Campbell91ASJ3] 39
 W. N. Campbell: “Foot-Level Timing Control for English Speech Systems,” 日本音響学会平成3年度春季研究発表会講演論文集, 1-6-13, pp. 231-232 (1991.03).

(W. N. Campbell: “Foot-level Timing Control for English Speech Systems,” *ASJ Spring Meeting, 1-6-3, pp. 231-232 (1991.03).*)

[Campbell91SP3] 41
 W. N. Campbell, Yoshinori Sagisaka: “音声タイミングにみられるモーラと音節の影響,” 電子情報通信学会技術研究報告, SP90-107, pp. 35-40 (1991.03).
 (W. N. Campbell, Yoshinori Sagisaka: “Moraic and Syllable-Level Effects on Speech Timing,” *IEICE Technical Report, SP90-107, pp. 35-40 (1991.03).*)

[Campbell91SL3] 47
 W. N. Campbell, 匂坂芳典: “日本語モーラタイミングについて,” 音声言語研究会資料 (1991.03).
 (W. N. Campbell, Yoshinori Sagisaka: “Mora-Level Effects in Japanese Speech Timing,” *Transaction on Spoken Language (1991.03).*)

[Campbell91ATR] 53
 W. N. Campbell: “Segmental Elasticity and Timing in Japanese Speech,” *Speech Perception, Production and Linguistic Structure*, pp. 403-418, eds. Y. Tohkura, Y. Sagisaka, E. V. Bateson, OHM Publishing, Japan (1991)

[Campbell91Eurospeech9] 69
 W. N. Campbell: “Phrase-Level Factors Affecting Timing In Speech,” Proc. Eurospeech’91, pp. 629-632, Genova, Italy (1991.09).

[Campbell91ICPhS8] 73
 W. N. Campbell: “Durational Shortening and Anaphoric Reference,” Proc. ICPhS’91, Vol. 2, pp. 286-289, Aix-en-Provence, France (1991.08).

[Campbell91PERILUS12] 77
 W. N. Campbell: “Prosodic Segmentation of Recorded Speech,” *PERILUS XIV*, pp. 101-106, Stockholm (1991.12).

◆

[Dantsuji91ICPhS8] 83
 Masataka Dantsuji, Shigeki Sagayama: “A Study on Distinctive Features and Feature Hierarchies through Phoneme Environment Clustering (PEC),” Proc. ICPhS’91, Vol. 3, pp. 190-193, Aix-en-Provence, France (1991.08).

[Fujiwara91Korea7] 87

Shingo Fujiwara, Yasuhiro Komori, Masahide Sugiyama: "A Hybrid System for Phoneme Labelling based on HMM and Spectrogram Reading Knowledge," 電子情報通信学会技術研究報告, SP91-38, pp. 1-6 (1991.07).

(*IEICE Technical Report; Proc. of Korea-Japan Joint Workshop on Advanced Technology of Speech Recognition and Synthesis*, pp. 121-126, Seoul (1991.07).)

[Fujiwara91ASJ3] 99

藤原 紳吾, 小森 康弘, 杉山 雅英: "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," 日本音響学会平成 3 年度春季研究発表会講演論文集, 2-5-16, pp. 85-86 (1991.03).

(*Shingo Fujiwara, Yasuhiro Komori, Masahide Sugiyama: "A Hybrid Automatic Phoneme Labelling System Based on HMM and Spectrogram Reading Knowledge (1)," ASJ Spring Meeting, 2-5-16, pp. 85-86 (1991.03).*)

[Fujiwara91ASJ10] 101

藤原 紳吾, 岩橋 直人, 小森 康弘, 杉山 雅英: "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," 日本音響学会平成 3 年度秋季研究発表会講演論文集, 2-5-20, pp. 87-88 (1991.10).

(*Shingo Fujiwara, Naoto Iwahashi, Yasuhiro Komori Masahide Sugiyama: "Hybrid Automatic Phoneme Labelling System Based on HMM and Spectrogram Reading Knowledge (2)," ASJ Fall Meeting, 2-5-20, pp. 87-88 (1991.10).*)

[Fukuzawa91ASJ10a] 103

福沢 圭二, 小森 康弘, 沢井 秀文, 杉山 雅英: "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," 日本音響学会平成 3 年度秋季研究発表会講演論文集, 3-5-10, pp. 109-110 (1991.10).

(*Keiji Fukuzawa, Yasuhiro Komori, Hidefumi Sawai, Masahide Sugiyama: "A Segment-based Speaker Adaptation Neural Network Applied to Continuous Speech Recognition," ASJ Fall Meeting, 3-5-10, pp. 109-110 (1991.10).*)

[Fukuzawa91ASJ10b] 105

福沢 圭二, 杉山 雅英: "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," 日本音響学会平成 3 年度秋季研究発表会講演論文集, 3-5-11, pp. 111-112 (1991.10).

(*Keiji Fukuzawa, Masahide Sugiyama: "Evaluation of Unsupervised Speaker Adaptation Using A Neural Network," 3-5-11, pp. 111-112 (1991.10).*)

[Fukuzawa91NNSP9] 107

Keiji Fukuzawa, Hidefumi Sawai, Masahide Sugiyama: "Segment-Based Speaker Adaptation by Neural Network," Proc. of Neural Networks for Signal Processings, pp. 442-451 (1991.10).

[Hattori91ASJ3] 117

服部 浩明, 嵯峨山 茂樹: "少量データを用いたコードブックマッピングによる話者適応化," 日本音響学会平成 3 年度春季研究発表会講演論文集, 1-5-23, pp. 49-50 (1991.03).

(*Hiroaki Hattori, Shigeki Sagayama: "Speaker Adaptation with Small Size Data Based on Codebook Mapping Algorithm," ASJ Spring Meeting, 1-5-23, pp. 49-50 (1991.03).*)

[Hattori91ICASSP5] 119

Hiroaki Hattori: "Speaker Adaptation Based on Markov Modeling of Speakers in Speaker-Independent Speech Recognition," Proc. of 1991 International Conference on Acoustics, Speech, and Signal Processing, S13.4, pp. 845-848, Toronto, Canada (1991.05).

[Hattori91SP12] 123

服部 浩明: "ニューラルネットワークを用いたテキスト独立な話者認識," 電子情報通信学会技術研究報告, SP91-90, pp. 71-78 (1992.12).

(*H. Hattori: "Text-Independent Speaker Recognition Using Neural Networks," IEICE Technical Report, SP91-90, pp. 71-78 (1992.03).*)

[Hedelin90JSC9] 131

Per Hedelin, Dieter Huber: "The CTH Speech Database: An Integrated Multilevel Approach," Journal Speech Communication, pp. 365-374, North-Holland (1990.09).

[Huber90SST11] 141

Dieter Huber: "Speech Style Variations of F_0 in a Cross-Linguistics Perspective," Proc. of SST'90 (1990.11).

[Huber90EDR11] 147

Dieter Huber: "An Electronic Dictionary for Computer Speech Applications," The EDR International Workshop on Electronic Dictionaries, pp. 1-11, Kanagawa (1990.11).

- [Huber91ICPhS8]..... 158
 Dieter Huber: "On the Discourse Function of Intonation," Proc. ICPhS'91, Vol. 5, pp. 190-193, Aix-en-Provence, France (1991.08).
- ◆
- [Iwahashi91ASJ3]..... 162
 岩橋 直人, 海木 延佳, 嵯峨山 茂樹, 匂坂 芳典: "音響的尺度を用いた合成単位素片の選抜法," 日本音響学会平成3年度春季研究発表会講演論文集, 1-6-1, pp. 207-208 (1991.03).
 (Naoto Iwahashi, Nobuyoshi Kaiki, Shigeki Sagayama, Yoshinori Sagisaka: "Unit Selection with Acoustic Measure for Speech Synthesis," ASJ Spring Meeting, 1-6-1, pp. 207-208 (1991.03).)
- [Iwahashi91SP5]..... 164
 岩橋 直人, 海木 延佳, 匂坂 芳典: "音響的尺度に基づく複合音声単位選抜法," 電子情報通信学会技術研究報告, SP91-5, pp. 33-40 (1991.05).
 (Naoto Iwahashi, Nobuyoshi Kaiki, Yoshinori Sagisaka: "Non-uniform Unit Selection based on Several Acoustic Measures for Speech Synthesis," IEICE Technical Report, SP91-5, pp. 33-40 (1991.05).)
- [Iwahashi91ASJ10]..... 172
 岩橋 直人, 藤原 紳悟, 小森 康弘, 杉山 雅英, 匂坂 芳典: "自動セグメンテーションによる音声合成単位の作成," 日本音響学会平成3年度秋季研究発表会講演論文集, 1-6-21, pp. 231-232 (1991.10).
 (Naoto Iwahashi, Shingo Fujiwara, Yasuhiro Komori, Masahide Sugiyama, Yoshinori Sagisaka: "Generation of Speech Synthesis Units by Automatic Labelling," ASJ Fall Meeting, 1-6-21, pp. 231-232 (1991.10).)
- ◆
- [Kaiki91ASJ3a]..... 174
 海木 延佳, W. N. Campbell, 匂坂 芳典: "制御構造を考慮した音韻継続時間長の設定," 日本音響学会平成3年度春季研究発表会講演論文集, 1-6-9, pp. 223-224 (1991.03).
 (Nobuyoshi Kaiki, W. N. Campbell, Yoshinori Sagisaka: "Incorporating Segmental Duration Setting from the Mora Level Control," ASJ Spring Meeting, 1-6-9, pp. 223-224 (1991.03).)
- [Kaiki91ASJ3b]..... 176
 海木 延佳, 匂坂 芳典: "種々の発話様式における音韻継続時間長の分析," 日本音響学会平成3年度春季研究発表会講演論文集, 1-6-10, pp. 225-226 (1991.03).

- (Nobuyoshi Kaiki, Yoshinori Sagisaka: "Analysis of Segmental Duration in Japanese Speech with Various Speaking Styles," ASJ Spring Meeting, 1-6-10, pp. 225-226 (1991.03).)
- [Kaiki91Eurospeech9]..... 178
 Nobuyoshi Kaiki, Katsuhiko Mimura, Yoshinori Sagisaka: "Statistical Modeling of Segmental Duration and Power Control for Japanese," Proc. Eurospeech'91, pp. 625-628, Genova, Italy (1991.09).
- [Kaiki91ASJ10a]..... 182
 海木 延佳, 匂坂 芳典: "音韻継続時間長制御における個人性の分析," 日本音響学会平成3年度秋季研究発表会講演論文集, 2-6-18, pp. 273-274 (1991.10).
 (Nobuyoshi Kaiki, Yoshinori Sagisaka: "Analysis of Personality for Segmental Duration Control in Japanese Speech," ASJ Fall Meeting, 2-6-18, pp. 273-274 (1991.10).)
- [Kaiki91ASJ10b]..... 184
 海木 延佳, 匂坂 芳典: "文音声におけるポーズ長制御の検討," 日本音響学会平成3年度秋季研究発表会講演論文集, 3-6-4, pp. 287-288, (1991.10).
 (Nobuyoshi Kaiki, Yoshinori Sagisaka: "A Study of Pause Duration Control in Japanese Sentence Speech," ASJ Fall Meeting, 3-6-4, pp. 287-288 (1991.10).)
- ◆
- [Katagishi91ASJ10]..... 186
 片岸 一起, 鷹見 淳一, 嵯峨山 茂樹: "HMM音素モデルにおける音声データの挿入及びシフト操作の効果," 日本音響学会平成3年度秋季研究発表会講演論文集, 2-5-19, pp. 85-86 (1991.10).
 (Kazuki Katagishi, Jun-ichi Takami, Shigeki Sagayama: "The Effects of Interpolation and Shift Operation for Labelled Speech Data on HMM Phoneme Models," ASJ Fall Meeting, 2-5-19, pp. 85-86 (1991.10).)
- ◆
- [Kato91ASJ10]..... 188
 加藤 宏明, 津崎 実, 匂坂 芳典: "単語内音韻継続時間長の弁別閾 - モーラ位置、母音種類、 F_0 輪郭の影響 -," 日本音響学会平成3年度秋季研究発表会講演論文集, 2-7-21, pp. 391-392 (1991.10).
 (Hiroaki Kato, Minoru Tsuzaki, Yoshinori Sagisaka: "Discrimination Threshold for Segmental Duration in Words - Effects of Mora Position, F_0 Contour, and Vowel Color -," ASJ Fall Meeting, 2-7-21, pp. 391-392 (1991.10).)
- [Kato91H11]..... 190

- 加藤 宏明, 津崎 実, 勾坂 芳典: “単語内音韻継続時間長の弁別閾 — モーラ位置、母音種類、 F_0 輪郭の影響 —,” 日本音響学会聴覚研究会資料, H-91-43, pp. 1-7 (1991.10).
(Hiroaki Kato, Minoru Tsuzaki, Yoshinori Sagisaka: “Discrimination Threshold for Segmental Duration in Words - Effects of Mora Position, F_0 Contour, and Vowel Color -,” Trans. ASJ Auditory Perception Committee, H-91-43, pp. 1-7 (1991.11).)
- ◆
- [Kita91ASJ10].....197
北 研二, 森元 暹, 嵯峨山 茂樹: “上位文法カテゴリへの到達可能性照合機構を備えた LR 解析法とその音声認識への応用,” 日本音響学会平成 3 年度秋季研究発表会講演論文集, 2-P-17, pp. 171-172 (1991.10).
(Kenji Kita, Tsuyoshi Morimoto, Shigeki Sagayama: “LR Parsing with a Category Reachability Test Applied to Speech Recognition,” ASJ Fall Meeting, 2-P-17, pp. 171-172 (1991.10).)
- ◆
- [Komori91ASJ3a].....199
小森 康弘, A. H. Waibel, 嵯峨山 茂樹: “音素識別ニューラルネットにおけるファジー学習法,” 日本音響学会平成 3 年度春季研究発表会講演論文集, 1-5-15, pp. 33-34 (1991.03).
(Yasuhiro Komori, A. H. Waibel, Shigeki Sagayama: “A New Fuzzy Training Method for Phoneme Identification Neural Networks,” ASJ Spring Meeting, 1-5-15, pp. 33-34 (1991.03).)
- [Komori91ASJ3b].....201
小森 康弘, 畑崎 香一郎: “特徴ベースによる音素セグメンテーションのロバスト性,” 日本音響学会平成 3 年度春季研究発表会講演論文集, 2-5-17, pp. 87-88 (1991.03).
(Yanesuhiro Komori, Kaichiro Hatazaki: “Robustness of a Feature Based Phoneme Segmentation,” ASJ Spring Meeting, 2-5-17, pp. 87-88 (1991.03).)
- [Komori91ICASSP5a].....203
Yasuhiro Komori, Kaichiro Hatazaki: “Robustness of a Feature Based Phoneme Segmentation System to Speaker Independent and Continuous Speech,” Proc. of 1991 International Conference on Acoustics, Speech, and Signal Processing, S14.2, pp. 885-888, Toronto, Canada (1991.05).
- [Komori91ICASSP5b].....207

- Yasuhiro Komori: “Time-State Neural Networks (TSNN) for Phoneme Identification by Considering Temporal Structure of Phonemic Features,” Proc. of 1991 International Conference on Acoustics, Speech, and Signal Processing, S2.22, pp. 125-128, Toronto, Canada (1991.05).
- [Komori91SP6].....211
小森 康弘, Alex H. Waibel, 嵯峨山 茂樹: “ニューラル・ファジー学習による TDNN-LR 連続音声認識システムの性能向上,” 電子情報通信学会技術研究報告, SP91-24, pp. 49-56 (1991.6).
(Yasuhiro Komori, Alex H. Waibel, Shigeki Sagayama: “An Improvement of TDNN-LR Continuous Speech Recognition System Using A Neural Fuzzy Training Approach,” IEICE Technical Report, SP91-24, pp. 49-56 (1991.06).)
- [Komori91IEICEtrans7].....219
Yasuhiro Komori, Kaichiro Hatazaki: “An Integration of Knowledge and Neural Networks toward a Phoneme Typewriter without a Language Model,” Trans. of IEICE, Vol. E 74, No. 7, pp. 1797-1805 (1991.07).
- [Komori91Eurospeech9].....228
Yasuhiro Komori, Kaichiro Hatazaki: “An Integration of Knowledge and Neural Networks Toward a Phoneme Typewriter Without a Language Model,” Proc. Eurospeech'91, pp. 1423-1426, Genova, Italy (1991.09).
- [Komori91ASJ10].....232
小森 康弘, 福沢 圭二, 杉山 雅英, A. H. Waibel, 嵯峨山 茂樹: “ニューラルファジー学習法の連続音声認識における効果,” 日本音響学会平成 3 年度秋季研究発表会講演論文集, 2-5-11, pp. 69-70 (1991.10).
(Yasuhiro Komori, Keiji Fukuzawa, Masahide Sugiyama, A. H. Waibel, Shigeki Sagayama: “Effectiveness of the Neural Fuzzy Training Method for Continuous Speech Recognition,” 2-5-11, pp. 69-70 (1991.10).)
- ◆
- [Kurematsu91JSAI5].....234
樽松 明: “自動翻訳電話のための音声情報処理,” 人工知能学会誌, Vol. 6, No. 3 (1991.05).
(Akira Kurematsu: “Speech Information Processing for Automatic Telephone Interpretation,” Journal of Japanese Society for Artificial Intelligence, Vol. 6, No. 3 (1991.05).)
- [Kurematsu91Tsushinkogyo6].....241

- 樽松 明: “音声認識,” 通信工業, 6月号, pp. 14-19 (1991.06).
(Akira Kurematsu: “Speech Recognition,” *Tsushin Kogyo*, Vol. 6, pp. 14-19 (1991.06).)
- [Kurematsu91DenkiTsushin10] 247
樽松 明: “自動翻訳電話技術の動向,” 電気通信, Vol. 54, No. 538, pp. 62-74 (1991.10).
(Akira Kurematsu: “Overview of Automatic Telephone Interpretation Technology,” *Denki Tsushin*, Vol. 54, No. 538, pp. 62-74 (1991.10).)
- [Kurematsu91NCF10] 260
Akira Kurematsu: “A Perspective of Speech Translation Research,” Proc. of 1991 National Communication Forum (1991.10).
- ◆
- [Ladd91ICPhS8] 281
D. R. Ladd, W. N. Campbell: “Theories of Prosodic Structure: Evidence From Syllable Duration,” Proc. ICPhS'91, Vol. 2, Aix-en-Provence, France, pp. 290-293 (1991.08).
- ◆
- [Mimura91ASJ3] 285
三村 克彦, 匂坂 芳典: “音声パワーの動特性の分析と制御,” 日本音響学会平成3年度春季研究発表会講演論文集, 1-6-14, pp. 233-234 (1991.03).
(Katsuhiko Mimura, Yoshinori Sagisaka: “Analysis of Temporal Patterns of Speech Power,” *ASJ Spring Meeting*, 1-6-14, pp. 233-234 (1991.03).)
- [Mimura91SP5] 287
三村 克彦, 海木 延佳, 匂坂 芳典: “統計的手法を用いた音声パワーの分析と動特性の制御,” 電子情報通信学会技術研究報告, SP91-4, pp. 25-32 (1991.05).
(Katsuhiko Mimura, Nobuyoshi Kaiki, Yoshinori Sagisaka: “Analysis and Control of Temporal Patterns of Speech Power Using Statistical Methods,” *IEICE Technical Report*, SP91-4, pp. 25-32 (1991.05).)
- [Mimura91Korea7] 295
Katsuhiko Mimura, Nobuyoshi Kaiki, Yoshinori Sagisaka: “Statistically Derived Rules for Amplitude and Duration Control in Japanese Speech synthesis,” 電子情報通信学会技術研究報告, SP91-41, pp. 19-24, (1991.07).
(*IEICE Technical Report; Proc. of Korea-Japan Joint Workshop on Advanced Technology of Speech Recognition and Synthesis*, pp. 151-156, Seoul (1991.07).)
- ◆
- [Minami91ASJ10] 307
南 康浩, 沢井 秀文, 中川正雄: “TDNNの構造の違いによる特性の比較,” 日本音響学会平成3年度春季研究発表会講演論文集, スペシャルセッション [ニューラルネットによる音声情報処理], 1-5-5, pp. 13-14 (1991.03).
- ◆
- [Morimoto91IEICEtrans7] 309
Tsuyoshi Morimoto, Kiyohiro Shikano, Kiyoshi Kogure, Hitoshi Iida, Akira Kurematsu: “Integration of Speech Recognition and Language Processing in a Japanese to English Spoken Language Translation System,” *Trans of IEICE*, Vol. E 74, No. 7, pp. 1889-1896 (1991.07).
- ◆
- [Murakami91ASJ3] 317
村上 仁一, 坪井 俊明: “BIGRAMをもちいた音節HMMによる文節音声認識,” 日本音響学会平成3年度春季研究発表会講演論文集, 3-5-15, pp. 117-118 (1991.03).
(Jin-ichi Murakami, Toshiaki Tsuboi: “Phrase Recognition Based on Syllable HMM and Bigram,” *ASJ Spring Meeting*, 3-5-15, pp. 117-118 (1991.03).)
- [Murakami91SP9] 319
村上 仁一, 荒木 哲郎, 池原 悟, 嵯峨山 茂樹: “音声におけるアクセント情報の持つ情報量の考察,” 電子情報通信学会技術研究報告, SP91-50, pp. 13-19 (1991.09).
(Jin-ichi Murakami, Tetsuro Araki, Satoru Ikehara, Shigeki Sagayama: “A study of Information Measure of Accents,” *IEICE Technical Report*, SP91-50, pp. 13-19 (1991.09).)
- [Murakami91IPS10] 326
村上 仁一: “漢字かなのTRIGRAMをもちいたかな漢字変換方法,” 情報処理学会第43回全国大会講演論文集, 7H-3, pp. 287-288 (1991.10).
(Jin-ichi Murakami: “Kanji Trigram Applied to Kana-Kanji Conversion,” *Proc. of the 43th National Convention IPS Japan*, 7H-3, pp. 287-288 (1991.10).)
- [Murakami91ASJ10] 328
村上 仁一, 嵯峨山 茂樹: “自由発話の音声認識における問題点の考察,” 日本音響学会平成3年度秋季研究発表会講演論文集, 2-P-26, pp. 189-190 (1991.10).
(Jin-ichi Murakami, Shigeki Sagayama: “Discussion on Problems in Spontaneous Speech Recognition,” *ASJ Fall Meeting*, 2-P-26, pp. 189-190 (1991.10).)

[Murakami91SP12].....	350
村上 仁一, 嵯峨山 茂樹: “自由発話音声認識における音響的および言語的な問題点の検討,” 電子情報通信学会技術研究報告, SP91-100, pp. 71-78 (1991.12). (Jin-ichi Murakami, Toshiaki Tsuboi: “A Discussion of Acoustic and Linguistic Problems in Spontaneous Speech Recognition,” <i>IEICE Technical Report, SP91-100, pp. 71-78 (1991.12).</i>)	
[Nagai91ASJ3].....	338
永井 明人, 嵯峨山 茂樹, 北 研二: “音素コンテキスト依存型 LR テーブルの生成アルゴリズム,” 日本音響学会平成 3 年度春季研究発表会講演論文集, 3-5-2, pp. 91-92 (1991.03). (Akito Nagai, Shigeki Sagayama, Kenji Kita: “Algorithm for generating a Phoneme-context-dependent LR Parsing Table,” <i>ASJ Spring Meeting, 3-5-2, pp. 91-92 (1991.03).</i>)	
[Nagai91SP6].....	340
永井 明人, 嵯峨山 茂樹, 北 研二: “HMM-LR 連続音声認識における音素環境依存型パーザの実現アルゴリズム,” 電子情報通信学会技術研究報告, SP91-23, pp. 41-48 (1991.06). (Akito Nagai, Shigeki Sagayama, Kenji Kita: “Algorithm for Phoneme-environment-dependent Parser in Continuous Speech Recognition Using HMM-LR Method,” <i>IEICE Technical Report, SP91-23, pp. 41-48 (1991.06).</i>)	
[Nagai91Eurospeech9].....	348
Akihito Nagai, Shigeki Sagayama: “Phoneme-context-dependent LR Parsing Algorithms for HMM-based Continuous Speech Recognition,” <i>Proc. Eurospeech'91, 48.3, pp. 1397-1400, Genova, Italy (1991.09).</i>	
[Nagai91ASJ10a].....	352
永井 明人, 鷹見 淳一, 嵯峨山 茂樹: “環境依存連続 HMM を用いた HMM-LR 連続音声認識,” 日本音響学会平成 3 年度秋季研究発表会講演論文集, 1-5-20, pp. 39-40 (1991.10). (Akito Nagai, Jun-ichi Takami, Shigeki Sagayama: “HMM-LR Continuous Speech Recognition Using Allophonic Continuous HMMs,” <i>ASJ Fall Meeting, 1-5-20, pp. 39-40 (1991.10).</i>)	
[Nagai91ASJ10b].....	354
永井 明人, 北 研二, 花沢 利行, 鈴木 忠, 岩崎 知弘, 川端 豪, 中島 邦男, 鹿野 清宏, 森元 暲, 嵯峨山 茂樹, 樽松 明: “HMM-LR 連続音声認識装置の開発と性能評価,” 日本音響学会平成 3 年度秋季研究発表会講演論文集, 1-5-23, pp. 45-46 (1991.10).	

(Akito Nagai, Kenji Kita, Toshiyuki Hanazawa, Tadashi Suzuki, Tomohiro Iwasaki, Takeshi Kawabata, Kunio Nakajima, Kiyohiro Shikano, Tsuyoshi Morimoto, Shigeki Sagayama, Akira Kurematsu: “Hardware Development for HMM-LR Continuous Speech Recognition System,” *ASJ Fall Meeting, 1-5-23, pp. 45-46 (1991.10).*)

[Nakamura91ICASSP5].....	356
Masami Nakamura, Shin-ichi Tamura, Shigeki Sagayama: “Phoneme Recognition by Phoneme Filter Neural Networks,” <i>Proc. of 1991 International Conference on Acoustics, Speech, and Signal Processing, S2.12, pp. 85-88, Toronto, Canada (1991.05).</i>	
[Nakamura91IEICEtrans10].....	360
中村 悟, 沢井 秀文: “不特定話者音素認識のためのニューラルネットワークアーキテクチャの性能比較,” 電子情報通信学会論文誌 D-II, Vol. J74-D-II, No. 10, pp. 1361-1369 (1991.10). n (Satoru Nakamura, Hidefumi Sawai: “Comparison of Performance on Neural Network Architectures for Speaker-Independent Phoneme Recognition,” <i>Trans. of IEICE D-II, Vol. J74-D-II, No. 10, pp. 1361-1369 (1991.10).</i>)	
[Nakamura91ASJ10].....	369
中村 悟, 沢井 秀文, 杉山 雅英: “TDNN を用いた不特定話者音素認識,” 日本音響学会平成 3 年度秋季講演論文集, 2-5-8, pp. 63-64 (1991.10). (Satoru Nakamura, Hidefumi Sawai, Masahide Sugiyama: “Speaker Independent Phoneme Recognition Using Time-Delay Neural Networks,” <i>ASJ Fall Meeting, 2-5-8, pp. 63-64 (1991.10).</i>)	
[Nakamura91ASJ10].....	371
中村 哲, 濱口 清治, 永井 明人, 鹿野 清宏, 田中 厚夫, 嵯峨山 茂樹, 樽松 明: “ベクトル量子化に基づく話者適応化装置の開発と性能評価,” 日本音響学会平成 3 年度秋季研究発表会講演論文集, 3-5-6, pp. 101-102 (1991.10). (Satoshi Nakamura, Seiji Hamaguchi, Akito Nagai, Kiyohiro Shikano, Atsuo Tanaka, Shigeki Sagayama, Akira Kurematsu: “Development of XQ-based Speaker Adaptation System,” <i>ASJ Fall Meeting, 3-5-6, pp. 101-102 (1991.10).</i>)	
[Ohkura91ASJ3].....	373

- 大倉 計美, 杉山 雅英: “雑音環境下における HMM と TDNN の文節認識性能の評価,” 日本音響学会平成 3 年度春季研究発表会講演論文集, 2-5-9, pp. 71-72 (1991.03).
(*Kazumi Ohkura, Masahide Sugiyama: “Phrase Recognition Using HMM-LR and TDNN-LR under Noisy Environment,” ASJ Spring Meeting, 2-5-9, pp. 71-72 (1991.03).*)
- [Ohkura91ICASSP5].....375
Kazumi Ohkura, Masahide Sugiyama: “Speech Recognition in a Noisy Environment Using a Noise Reduction Neural Network and a Codebook Mapping Technique,” Proc. of 1991 International Conference on Acoustics, Speech, and Signal Processing, S14.13, pp. 929-932, Toronto, Canada (1991.05).
- [Ohkura91SP9].....379
大倉 計美, 杉山 雅英: “セグメント特徴量を用いた雑音環境下での HMM 音声認識,” 電子情報通信学会技術研究報告, SP91-55, pp. 1-6 (1991.9).
(*Kazumi Ohkura, Masahide Sugiyama: “Segment-based HMM Applied to Noisy Speech Recognition,” IEICE TEchnical Report, SP91-55, pp. 1-6 (1991.09).*)
- [Ohkura91ASJ10].....385
大倉 計美, 杉山 雅英: “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,” 日本音響学会平成 3 年度秋季研究発表会講演論文集, 1-5-6, pp. 11-12 (1991.10).
(*Kazumi Ohkura, Masahide Sugiyama: “Segment-based HMM Applied to Noisy Speech Recognition,” ASJ Fall Meeting, 1-5-6, pp. 11-12 (1991.10).*)
- ◆
- [Rainton91ASJ3].....387
David Rainton, 嵯峨山 茂樹: “Recognising English Speech Using the ATR HMM-LR,” 日本音響学会平成 3 年度春季研究発表会講演論文集, 3-5-17, pp. 121-122 (1991.03).
(*David Rainton, Shigeki Sagayama: “Recognising Speech Using the ATR HMM-LR,” ASJ Spring Meeting, 3-5-17, pp. 121-122 (1991.03).*)
- [Rainton91ICASSP5].....389
David Rainton: “Speech Recognition - A Time-Frequency Subspace Filtering Based Approach,” Proc. of 1991 International Conference on Acoustics, Speech, and Signal Processing, S7.13, pp. 461-464, Toronto, Canada (1991.05).
- [Rainton91SP9].....393
David Rainton, Shigeki Sagayama: “Class Specific Observation Vector Orthogonalisation and Its Interpretation as a Form of Tied Continuous Mixture HMM,” 電子情報通信学会技術研究報告, SP91-56, pp. 7-13 (1991.09).
(*David Rainton, Shigeki Sagayama: “Class Specific Observation Vector Orthogonalisation and Its Interpretation as a Form Of Tied Continuous Mixture HMM,” IEICE Technical Report, SP91-56, pp. 7-13 (1991.09).*)
- [Rainton91ASJ10].....400
David Rainton, 嵯峨山 茂樹: “A New Tied Continuous Mixture Density HMM via Orthogonalisation of the Full Covariance Observation Matrix,” 日本音響学会平成 3 年度秋季研究発表会講演論文集, 2-5-15, pp. 77-78 (1991.10).
(*David Rainton, Shigeki Sagayama: “A New Tied Continuous Mixture Density HMM Via Orthogonalisation of The Full covariance Observation Matrix,” ASJ Fall Meeting, 2-5-15, pp. 77-78 (1991.10).*)
- ◆
- [Sagayama91Book]..... (not included in this collection)
Shigeki Sagayama: “Speech Analysis,” in *Speech Science and Technology* (S. Saito ed.), pp. 64-96 and pp. 318-326, Ohmsha Pub., Tokyo (in Japan) and IOS Press (other countries) (1992).
- [Sagayama91SPREC2].....402
嵯峨山 茂樹: “HMM 尤度行列演算による音声認識アルゴリズム,” 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 (箱根強羅), pp. 39-42 (1991.02).
(*Shigeki Sagayama: “A Speech Recognition Algorithm Based on HMM Likelihood Matrix Computation,” SPREC Symposium (Hakone), pp. 39-42 (1991.02).*)
- [Sagayama91TohokuSympo2].....406
嵯峨山 茂樹: “数理統計モデルに基づく音声認識,” 第 27 回東北大通研シンポジウム「パターンの認識・理解における諸問題とその実現」予稿集, 2-6, pp. 119-127 (1991.02).
(*Shigeki Sagayama: “Speech Recognition Based on Statistical Models,” Proc. of 27th Symposium, Institute of Electrical Communications, Tohoku University, 2-6, pp. 119-127 (1991.02).*)
- [Sagayama91ISCIEcourse2].....415
嵯峨山 茂樹: “最近の音声認識研究,” 第 40 回システム制御情報講習会「ここまできた高度情報化技術」テキスト, システム制御情報学会, pp. 93-116 (1991.02).
(*Shigeki Sagayama: “Recent and Current Status of Speech Recognition Research,”*

[Sagayama91ICASSP5]..... 439
 Shigeki Sagayama, Jun-ichi Takami, S.Honma, "An Allophone Clustering Technique Applied to Large Vocabulary Word Speech Recognition," Proc. of 1991 International Conference on Acoustics, Speech, and Signal Processings, 56.S3.1 (not appeared in the proceedings), Toronto, Canada (1991.05).

[Sagayama91Korea7]..... 443
 Shigeki Sagayama, Tsuyoshi Morimoto, Akira Kurematsu: "Speech Recognition Toward an Interpreting Telephone," 電子情報通信学会技術研究報告, SP91-29, pp. 7-12 (1991.07);
 (*IEICE Technical Report; Proc. of Korea-Japan Joint Workshop on Advanced Technology of Speech Recognition and Synthesis, pp. 22-27, Seoul (1991.07),*)

[Sagayama91PatentAward6]..... 455
 嵯峨山 茂樹, 板倉 文忠: "複合正弦波音声合成法の発明," 全国発明表彰受賞者功績概要, 社団法人発明協会, pp. 30-31 (1991.06).
 (*Shigeki Sagayama, Fumitada Itakura: "Speech Synthesis by Composite Sinusoidal Model," National Convention Awards - Summary of Winning Entries, Japan Institute of Invention and Innovation, pp. 30-31 (1991.06).*)

[Sagayama91IEICEcourse7]..... 457
 嵯峨山 茂樹: "音声認識の基本単位," 信学会講習会「音声認識・合成技術の現状と将来課題」テキスト, pp. 7-14 (1991.07).
 (*Shigeki Sagayama: "Units for Speech Recognition," Tutorial Course of IEICEJ on the Current Status and Future Problems in Speech Recognition and Synthesis, pp. 7-14 (1991.07).*)

[Sagayama91Eurospeech9]..... 465
 Shigeki Sagayama: "A matrix representation of HMM-based speech recognition algorithms," Proc. of Eurospeech'91, 42.5, pp. 1225-1228, Genova, Italy (1991.09).

[Sagayama91IPScourse10]..... 469
 嵯峨山 茂樹: "音声とヒューマンインタフェース," 情報処理学会関西支部講習会 (1991年10月29-30日), pp. 6.1-22 (1991.10).
 (*Shigeki Sagayama, "Speech and Human Interface," Tutorial Course of Information Processing Society, pp. 6.1-22 (1991.10).*)

[Sagisaka91JASA4]..... 491
 Yoshinori Sagisaka: "The Analysis of F_0 Reset in Relation to Phrase Dependency Structure," The Journal of the Acoustical Society of America, Vol. 89, No. 4, Pt. 2, pp. 1937 (1991.04).

[Sagisaka91IEICJcourse7]..... 492
 匂坂 芳典: "音声合成の基本単位と最適化技術 [音声認識・合成技術の現状と将来課題]," 電子情報通信学会講習会 pp. 77-82 (1991.07).
 (*Yoshinori Sagisaka: "Speech Synthesis Unit and the Optimization of Control Rules," IEICE Tutorial Course of Information Processing Society, pp. 77-82 (1991.07).*)

[Sagisaka91ICPhS8]..... 498
 Yoshinori Sagisaka, Nobuyoshi Kaiki: "Prosody Control for Spontaneous Speech Synthesis," Proc. ICPhS'91, Vol. 3, pp. 506-509, Aix-en-Provence, France (1991.08).

[Sagisaka91Phonetics9]..... 502
 匂坂 芳典: "音声合成の立場から," 日本音声学会1991年全国大会シンポジウム予稿集, pp. 3-4 (1991.09).
 (*Yoshinori Sagisaka: "Future Issues for Speech Research ; Speech Synthesis," Proc. Japan Speech Society Symposium 1991, pp. 3-4 (1991.09).*)

[Sagisaka91ASJ10]..... 504
 匂坂 芳典: "会話音声に見られる音調バタンの特徴," 日本音響学会平成3年度秋季研究発表会講演論文集, 2-6-8, pp. 253-254 (1991.10)
 (*Yoshinori Sagisaka: "Characteristics of Fundamental Frequency Contour in Conversational Speech," ASJ Fall Meeting, 2-6-8, pp. 253-254 (1991.10).*)

[Sagisaka91sympo10]..... 506
 匂坂 芳典: "音声タイミング制御にみられる日本語の特徴," 第36回日本音声言語医学会シンポジウム予稿集, pp. 17-21 (1991.10).
 (*Yoshinori Sagisaka: "Japanese Temporal Organization Manifested in Segmental Duration Characteristics," pp. 17-21 (1991.10).*)

[Sawai91ASJ3]..... 511
 沢井 秀文, 中村 悟, 福沢 圭二, 杉山 雅英: "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," 日本音響学会平成3年度春季講演論文集,

- 1-5-17, pp. 37-38 (1991.03).
 (Hidefumi Sawai, Satoru Nakamura, Keiji Fukuzawa, Masahide Sugiyama: "On Connectionist Approaches to Speaker-independent Recognition," *ASJ Spring Meeting, 1-5-17, pp. 37-38 (1991.03).*)
- [Sawai91ICASSP5a]..... 513
 Hidefumi Sawai: "Frequency-Time-Shift-Invariant Time-Delay Neural Networks for Robust Continuous Speech Recognition," Proc. of 1991 International Conference on Acoustics, Speech, And Signal Processing, S2.1, pp. 45-48, Toronto, Canada (1991.05).
- [Sawai91ICASSP5b]..... 517
 Hidefumi Sawai: "TDNN-LR Continuous Speech Recognition System Using Adaptive Incremental TDNN Training," Proc. of 1991 International Conference on Acoustics, Speech, And Signal Processing, S2.4, pp. 53-56, Tronto, Canada (1991.05).
- [Sawai91IEICEtrans7]..... 521
 Hidefumi Sawai, Yasuhiro Minami, Masanori Miyatake, Alex Waibel, Kiyohiro Shikano: "Connectionist Approaches to Large Vocabulary Continuous Speech Recognition," Special Issue on Continuous Speech Recognition and Understanding, Trans. of IEICE, Vol.E 74, No.7, pp. 1834-1844 (1991.07).
- [Sawai91Eurospeech9]..... 532
 Hidefumi Sawai, Satoru Nakamura: "Time-Delay Neural Network Architectures for High-Performance Speaker-Independent Recognition," Proc. of EuroSpeech'91, pp. 1011-1014, Genova, Italy (1991.09).
- ◆
- [Shimodaira91ASJ3]..... 536
 下平 博, 木村 正行, 嵯峨山 茂樹: "ピッチパタンのクラスタリングによる連続音声の句境界検出," 日本音響学会平成3年度春季研究発表会講演論文集, 2-5-14, pp. 81-82 (1991.03).
 (Hiroshi Shimodaira, Masayuki Kimura, Shigeki Sagayama: "Phrase Boundary Detection Using Pitch Contour Clustering," *ASJ Spring Meeting, 2-5-14, pp. 81-82 (1991.10).*)
- ◆
- [Singer91SP9]..... 538
 Harald Singer, 嵯峨山 茂樹: "ピッチとスペクトルの相関を用いたHMM音素認識," 電子情報通信学会技術研究報告, SP91-57, pp. 15-20 (1991-09).
 (Harald Singer, Shigeki Sagayama: "Use of Correlation between Pitch and Spectral Parameters for HMM Phoneme Recognition," *IEICE Technical Report, SP91-57, pp. 15-20 (1991.09).*)
- [Singer91ASJ10]..... 544
 Harald Singer, Shigeki Sagayama: "Pitch-cepstrum Correlation Used for HMM Phoneme Recognition," 日本音響学会平成3年度秋季研究発表会講演論文集, 2-P-5, pp. 147-148 (1991.10).
 (Harald Singer, Shigeki Sagayama: "Pitch-cepstrum Correlation used for HMM phoneme Recognition," *ASJ Fall Meeting, 2-P-5, pp. 147-148 (1991.10).*)
- ◆
- [Sugiyama91ASJ3]..... 546
 杉山 雅英: "LPCケプストラム係数の存在領域について," 日本音響学会平成3年度春季研究発表会講演論文集, 2-6-4, pp. 253-254 (1991.03).
 (Masahide Sugiyama: "Stable Space for LPC Cepstrum Coefficients," *ASJ Spring Meeting, 2-6-4, pp. 253-254 (1991.03).*)
- [Sugiyama91IEICEtrans3]..... 548
 Masahide Sugiyama: "Distortion Geodesic Lines and Their Application to Spectral Interpolation," Trans of IEICE, Vol.E-74, No.3, pp. 609-614 (1991.03).
- [Sugiyama91ICASSP5]..... 554
 Masahide Sugiyama: "Automatic Language Recognition Using Acoustic Features," Proc. of 1991 International Conference on Acoustics, Speech, And Signal Processing, S12.8, pp. 813-816, Tronto, Canada (1991.05).
- [Sugiyama91ISCAS6]..... 558
 Masahide Sugiyama: "Review of TDNN (Time Delay Neural Network) Architectures for Speech Recognition," Proc. ISCAS'91, Singapore (1991.06).
- [Sugiyama91ASJ10a]..... 562
 杉山 雅英: "Bilingual 音声を対象とした日英言語識別の検討," 日本音響学会平成3年度秋季研究発表会講演論文集, 3-6-12, pp. 133-134 (1991.10).
 (Masahide Sugiyama: "Language Identification for Bilingual Speech," *ASJ Fall Meeting, 3-6-12, pp. 133-134 (1991.10).*)
- [Sugiyama91ASJ10b]..... 564

杉山 雅英, Alain Biem: "TDNN-HMM による音素認識," 日本音響学会平成 3 年度秋季研究発表会講演論文集, 2-P-6, pp. 149-150 (1991.10)
(Masahide Sugiyama, Alain Biem: "TDNN-HMM Approach to Phoneme Recognition," ASJ Fall Meeting, 2-P-6, pp. 149-150 (1991.10).)

[Takami91SPREC2].....566

鷹見 淳一, 嵯峨山 茂樹, "分類型ニューラルネットワークを用いたおけるロバストネスの問題について," 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 (箱根強羅), pp. 57-58 (1991.02).
(Jun-ichi Takami and Shigeki Sagayama: "On Robustness of Classification-type Neural Networks," SPREC Symposium (Hakone), pp. 57-58 (1991.02).)

[Takami91ASJ3].....568

鷹見 淳一, 片岸 一起, 嵯峨山 茂樹: "音素環境情報を利用した連続型 HMM による音素認識," 日本音響学会平成 3 年度春季研究発表会講演論文集, 3-P-12, pp. 171-172 (1991.03).
(Jun-ichi Takami, Kazuki Katagishi, Shigeki Sagayama: "Phoneme Recognition by Continuous HMM with Phoneme Environment Information," ASJ Spring Meeting, 3-P-12, pp. 171-172 (1991.03).)

[Takami91ICASSP5].....570

Jun-ichi Takami, Shigeki Sagayama: "A Pairwise Discriminant Approach to Robust Phoneme Recognition by Time-Delay Neural Networks," Proc. of 1991 International Conference on Acoustics, Speech, and Signal Processing, S2.13, pp. 89-92, Toronto, Canada (1991.05).

[Takami91SP6].....574

鷹見 淳一, 嵯峨山 茂樹: "単一ガウス分布 HMM の音素環境木構造に基づく平滑化による頑健な音素認識," 電子情報通信学会技術研究報告, SP91-19, pp. 9-16 (1991.06).
(Jun-ichi Takami, Shigeki Sagayama: "Robust Phoneme Recognition using Single Gaussian Density HMMs with Smoothing Based on the Structure of the Phoneme Environment Tree," IEICE Technical Report, SP91-19, pp. 9-16 (1991.06).)

[Takami91NNSP9].....582

Jun-ichi Takami, Atsuhiko Kai, Shigeki Sagayama: "Speech Recognition by Combining Pairwise Discriminant Time-Delay Neural Networks and Predictive LR-Parser," Proc. of Neural Networks for Signal Processings, pp. 327-336, Princeton (1991.09).

[Takami91ASJ10a].....592

鷹見 淳一, 甲斐 充彦, 嵯峨山 茂樹: "対判定型 TDNN と予測 LR パーザの結合による連続音声認識," 日本音響学会平成 3 年度秋季研究発表会講演論文集, 1-5-22, pp. 43-44 (1991.10).

(Jun-ichi Takami, Atsuhiko Kai, Shigeki Sagayama: "Continuous Speech Recognition by Combining PD-TDNNs and Predictive LR Parser," ASJ Fall Meeting, 1-5-22, pp. 43-44 (1991.10).)

[Takami91ASJ10b].....594

鷹見 淳一, 嵯峨山 茂樹: "逐次状態分割法 (SSS) による隠れマルコフネットワークの自動生成," 日本音響学会平成 3 年度秋季研究発表会講演論文集, 2-5-13, pp. 73-74 (1991.10).

(Jun-ichi Takami, Shigeki Sagayama: "Automatic Generation of the Hidden Markov Network by the Successive State Splitting Algorithm," ASJ Fall Meeting, 2-5-13, pp. 73-74 (1991.10).)

[Takami91SP12].....596

鷹見 淳一, 嵯峨山 茂樹: "音素コンテキストと時間に関する逐次状態分割による隠れマルコフ網の自動生成," 電子情報通信学会技術研究報告, SP91-88, pp. 57-64 (1991.12).

(Jun-ichi Takami, Shigeki Sagayama: "Automatic Generation of the Hidden Markov Network by Successive State Splitting on Contextual Domain and Temporal Domain," IEICE Technical Report, SP91-88, pp. 57-64 (1991.12).)

[Takeda91JASA12].....604

Kazuya Takeda, Yoshinori Sagisaka, Hisao Kuwabara: "On Sentence-level Factors Governing Segmental Duration in Japanese," The TeX Journal of the Acoustical Society of America, Vol.86(6), pp. 2081-2087 (1989.12).

[Takezawa91IEICE9].....611

竹澤 寿幸, 森元 逞, 樽松 明: "日英音声言語翻訳実験システム SL-Trans2 における音声対訳処理," 電子通信学会秋季全国大会講演論文集, D-53, 分冊 6, pp. 53-54 (1991.09).

(Toshiyuki Takezawa, Tsuyoshi Morimoto, Akira Kurematsu: "Interaction Processing in SL-TRANS 2: an Experimental System for Translating Japanese Speech to English," Proc. of the Institute of Electronics and Communication Engineers of Japan, Vol. 6, pp. 53-54 (1991.09).)

[Takezawa91ASJ10].....613

竹沢 寿幸, 大倉 計美, 森元 暉, 嵯峨山 茂樹, 樽松 明: “日英音声言語翻訳実験システム SL-TRANS2,” 日本音響学会平成3年度秋季研究発表会講演論文集, 1-5-24, pp. 47-48 (1991.10).

(Toshiyuki Takezawa, Kazumi Ohkura, Tsuyoshi Morimoto, Shigeki Sagayama, Akira Kurematsu: “SL-TRANS 2: an Experimental System for Translating Japanese Speech to English,” ASJ Fall Meeting, 1-5-24, pp. 47-48 (1991.10).)



[Umezaki91IEICEtrans4]..... 615

梅崎 太造, Harald Singer, 板倉 文忠: “平滑化群遅延スペクトル距離尺度の不特定話者音声認識における評価,” 電子情報通信学会論文誌 A, Vol. J74-A, No. 4, pp. 610-618 (1991.04).

(Taizo Umezaki, Harald Singer, Fumitada Itakura: “Evaluation of the Smoothed Group Delay Spectrum Distance Measure in Speaker-Independent Speech Recognition,” Trans. of IEICE, Vol. J74-A, No. 4, pp. 610-618 (1991.04).)

[Umezaki91ROC11]..... 624

Taizo Umezaki, Harald Singer, Fumitada Itakura: “Evaluation of the Smoothed Group Delay Spectrum Distance Measure in Speaker-Independent Speech Recognition,” 1. ROC-Japan Seminar on New Speech Recognition Method, pp. 179-188, National Chiao Tung University, Taiwan (1991.11).

List of Publications, January through December, 1992

Araki92IPS03 1
 荒木 哲郎, 池原 悟, 村上 仁一: “音節ラティスに適用するビタービアルゴリズムの評価について,” 情報処理学会第44回全国大会講演論文集, 7N-2, pp. 2.163-164 (1992.03).
 T. Araki, S. Ikehara, J. Murakami: “Evaluation of Viterbi Algorithm Applied for Syllable Lattices,” Proc. of the 44th National Convention IPS Japan, 7N-2, pp. 2.163-164 (1992.03).

◆

Biem92ACRCS10 3
 Alain Biem, Masahide 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 (1992.10).

◆

Campbell92ASJ03 16
 Wilhelm N. Campbell: “Gamma Modelling of English Segmental Durations,” 日本音響学会平成4年度春季研究発表会講演論文集, 1-2-18, pp. 241-242 (1992.03).
 ウィルヘルム N. キャンベル: “ガンマ分布を用いた英語音韻時間長のモデリング,” ASJ Spring Meeting, 1-2-18, pp. 241-242 (1992.03).

Campbell92ASJ03 18
 Wilhelm N. Campbell: “Prosodic Phrasing from Normalised Acoustic Measures,” 日本音響学会平成4年度春季研究発表会講演論文集, 1-2-19, pp. 243-244 (1992.03).
 ウィルヘルム N. キャンベル: “正規化した音響的尺度を用いた韻律フレーズング,” ASJ Spring Meeting, 1-2-19, pp. 243-244 (1992.03).

Campbell92SP03 20
 W. N. Campbell: “Synthesis Units for Natural English Speech” IEICE Technical Report, SP91-129, pp. 55-62 (1992.03).
 ウィルヘルム N. キャンベル: “自然な英語音声合成のための音声単位” 電子情報通信学会技術研究報告, SP91-129, pp. 55-62 (1992.03).

Campbell92ICSLP10a 28

Wilhelm N. Campbell: “Prosodic Encoding of English Speech” Proc. of International Conference on Spoken Language Processing, Th.fPM.3.3, pp. 663-666, Canada (1992.10).

Campbell92ICSLP10b 32
 W. N. Campbell & C. W. Wightman: “Prosodic Coding of Syntactic Structure for Speech Synthesis,” Proc. of International Conference on Spoken Language Processing, Th.fAM.2.4, pp. 1167-1170, Canada (1992.10).

Campbell92ASJ10 36
 W. N. Campbell: “Labelling an English Speech Database for Prosody Control,” ASJ Fall Meeting, 1-P-8, pp. 315-316 (1992.10).

Campbell92SST12 38
 W. N. Campbell, Y. Sagisaka: “Automatic Annotation of Speech Corpora,” Proc. Fourth Australian International Conference on Speech Science and Technology, pp. 686-691 (1992.12).

Campbell92narakokusai 44
 W. N. Campbell: “Speech Timing in English and Japanese,” Mom-busho International Symposium on Japanese Prosody, pp. 207-216, Nara Japan (1992).

Campbell92Book 54
 W. N. Campbell: “Syllable-based Segmental Duration,” Talking Machines: Theories, Models & Applications, Elsevier-North-Holland Pub. pp. 211-224 (1992).

◆

Fujiwara92ISSPA08 68
 Shingo Fujiwara, Yasuhiro Komori, Masahide Sugiyama: “An Integrated System for Automatic Labelling Based on HMM and Spectrogram Reading Knowledge,” Proc. of International Symposium on Signal Processing and Its Applications, pp. 275-278, Australia (1992.08).

110

Fujiwara92ICSLP10.....	72
Shingo Fujiwara, Yasuhiro Komori, Masahide Sugiyama: "A Phoneme Labelling Workbench using HMM and Spectrogram Reading Knowledge," Proc. of International Conference on Spoken Language Processing, Vol. 1, pp. 791-794 (1992.10).	
Fujiwara92ASJ10.....	76
藤原 紳吾, 小森 康弘, 杉山 雅英: "音素ラベリングワークベンチ," 日本音響学会平成4年度秋季研究発表会講演論文集, 2-Q-28, pp. 227-228 (1992.10). S.Fujiwara, Y.Komori, M.Sugiyama: "A Phoneme Labelling Workbench," ASJ Fall Meeting, 2-Q-28, pp. 277-278 (1992.10).	
◆	
Fukuzawa92SP01.....	78
福沢 圭二, 小森 康弘, 沢井 秀文, 杉山 雅英, "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," 電子情報通信学会技術研究報告, SP91-105, pp. 23-29 (1992.01). Keiji Fukuzawa, Yasuhiro Komori, Hidehumi Sawai, Masahide Sugiyama: "A Segment-based Speaker Adaptation Neural Network applied to TDNN-LR Continuous Speech Recognizer," IEICE Technical Report, SP91-105, pp. 23-29 (1992.01).	
Fukuzawa92ASJ03.....	85
福沢 圭二, 小森 康弘, 杉山 雅英: "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比較," 日本音響学会平成4年度春季研究発表会講演論文集, 2-Q-21, pp. 199-200 (1992.03). Keiji Fukuzawa, Yasuhiro Komori, Masahide Sugiyama: "A Comparison between Multi-Speaker Trained TDNN and Speaker Adaptation Neural Network in a TDNN-LR Continuous Speech Recognition System," ASJ Spring Meeting, 2-Q-21, pp. 199-200 (1992.03).	
Fukuzawa92ICASSP03.....	87
Keiji Fukuzawa, Yasuhiro Komori, Hidehumi Sawai, Masahide Sugiyama: "A Segment-based Speaker Adaptation Neural Network Applied to Continuous Speech Recognition," Proc. of 1992 International Conference on Acoustics, Speech, and Signal Processing, 55.1, Vol. 1, pp. 433-436 (1992.03).	
Fukuzawa92IEICE09.....	91

福沢 圭二, 杉山 雅英: "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," 電子情報通信学会秋季大会講演論文集, Vol. 1, pp. 271-272 (1992.09).	
Keiji Fukuzawa, Masahide Sugiyama: "Unsupervised Speaker Adaptation Using A Hierarchical Clustering Method and A Neural Network," Proc. of the 1992 IEICE Fall Conference, Vol. 1, pp. 271-272 (1992.09).	
Fukuzawa92ASJ10.....	93
福沢 圭二, 加藤 喜永, 杉山 雅英: "FPM-LR を用いた不特定話者連続音声認識," 日本音響学会平成4年度秋季研究発表会講演論文集, 3-7-3, pp. 167-168 (1992.10).	
Keiji Fukuzawa, Yoshinaga Kato, Masahide Sugiyama: "Speaker-Independent Continuous Speech Recognition using A FPM-LR System," ASJ Fall Meeting, 3-7-3, pp. 167-168 (1992.10).	
Fukuzawa92ICSLP10.....	95
Keiji Fukuzawa, Yoshinaga Kato and Masahide Sugiyama: "A Fuzzy Partition Model(FPM)Neural Network Architecture for Speaker-independent Continuous Speech Recognition," Proc. of 1992 International Conference on Spoken Language Processing, Vol. 2, pp. 1383-1386 (1992.10).	
Fukuzawa92SP12.....	99
福沢 圭二, 加藤 喜永, 杉山 雅英: "FPM-LR を用いた不特定話者連続音声認識の実現," 電子情報通信学会技術研究報告, SP92-107, pp. 31-38 (1992.12). Keiji Fukuzawa, Yoshinaga Kato, Masahide Sugiyama: "Speaker-Independent Continuous Speech Recognition Using FPM-LRs," IEICE Technical Report, SP92-107, pp. 31-38, (1992.12).	
◆	
Gurgen92IECEtrans01.....	107
Fikret S. Gurgen, Shigeki Sagayama, Sadaoki Furui, "A Study of Line Spectrum Pair Frequency Representation for Speech Recognition," IEICE Trans. Fundamentals, Vol. E75-A, No. 1, pp. 98-102 (1992.01).	
◆	
Hattori92ASJ03.....	112
服部 浩明, 嵯峨山 茂樹: "少数語彙による移動ベクトル場平滑化話者適応方式の文節認識による評価," 日本音響学会平成4年度春秋研究発表会講演論文集, 2-Q-15,	

pp. 187-188 (1992.03).
 H. Hattori, S. Sagayama: "Evaluation of Transfer Vector Field Smoothing Speaker Adaptation Method on Japanese Phrase Recognition," ASJ Spring Meeting, 2-Q-15, pp. 187-188 (1991.03).

Hattori92ICASSP03 114
 Hiroaki Hattori: "Tex-independent Speaker Recognition Using Neural Networks," Proc. of 1992 International Conference on Acoustics, Speech and Signal Processing, pp. II-153-156, San Francisco (1992.03).

Hattori92SP06 118
 服部 浩明, 嵯峨山 茂樹: "移動ベクトル平滑化話者適応の原理とアルゴリズム," 電子情報通信学会技術研究報告, SP92-15, pp. 15-22 (1992.06).
 H. Hattori, S.Sagayama: "Speaker Adaptation Based on Vector Field Smoothing," IEICE Technical Report, SP92-15, pp. 15-22 (1992.06).

Hattori92ICSLP10 126
 H. Hattori and S.Sagayama: "Vector Field Smoothing Principle For Speaker Adaptation," Proc. of International Conference on Spoken Language Processing, Vol. 1, pp. 381-384 (1992.10).

◆

Iwahashi92ICASSP03 130
 Naoto Iwahashi, Nobuyoshi Kaiki, Yoshinori Sagisaka: "Concatenative Speech Synthesis by Minimum Distortion Criteria," Proc. of 1992 International Conference on Acoustics, Speech, and Signal Processing, Vol. II, pp. 65-68, (1992.03).

Iwahashi92ASJ03 134
 岩橋 直人, 匂坂 芳典: "接続歪みを最小化する音声素片セットの構成法," 日本音響学会平成4年度春季研究発表会講演論文集, 1-2-6, pp. 217-218, (1992.3).
 Naoto Iwahashi, Yoshinori Sagisaka: "Organization Method of Speech Synthesis Segment Set Minimizing Concatenative Distortion," ASJ Spring Meeting, 1-2-6, pp. 217-218, (1992.03).

Iwahashi92ASJ03 136
 岩橋 直人, 匂坂 芳典: "歪み最小化音声合成手法の主観・客観評価," 日本音響学会平成4年度春季研究発表会講演論文集, 2-2-15, pp. 281-282, (1992.03).

Naoto Iwahashi, Yoshinori Sagisaka: "Objective and Subjective Evaluation of Segment Selection Method by Minimization of Acoustic Distortion," ASJ Spring Meeting, 2-2-15, pp. 281-282, (1992.03)

Iwahashi92SP09 138
 岩橋 直人, 匂坂 芳典: "音声素片ネットワーク最適化による合成素片セットの構成法," 電子情報通信学会技術研究報告, SP92-51, pp. 9-16, (1992.09).
 Naoto Iwahashi, Yoshinori Sagisaka: "Speech Segment Network Approach for an Optimal Synthesis Unit Set," IEICE Technical Report, SP92-51, pp. 9-16, (1992.09).

Iwahashi92ICSLP10 146
 Naoto Iwahashi, Yoshinori Sagisaka: "Speech Segment Network Approach for an Optimal Synthesis Unit Set," Proc. of 1992 International Conference on Spoken Language Processing, pp. 479-482, (1992.10).

Iwahashi92ASJ10a 150
 岩橋 直人, 匂坂 芳典: "空間分割型数量化法による音声制御の統計モデリング," 日本音響学会平成4年度秋季研究発表会講演論文集, 1-5-11, pp. 237-238, (1992.10).
 Naoto Iwahashi, Yoshinori Sagisaka: "Statistical Modeling of Speech by Space Partition Regression," ASJ Fall Meeting, 1-5-11, pp. 237-238, (1992.10).

Iwahashi92ASJ10b 152
 岩橋 直人, 匂坂 芳典: "音声素片ネットワーク最適化法による DIPHONE 素片セットの構成," 日本音響学会平成4年度秋季研究発表会講演論文集, 1-P-18, pp. 335-336, (1992.10).
 Naoto Iwahashi, Yoshinori Sagisaka: "Construction of DIPHONE Segment Set by Speech Segment Network Approach," ASJ Fall Meeting, 1-P-18, pp. 335-336, (1992.10).

◆

Kaiki92ASJ03a 154
 海木 延佳, 匂坂 芳典: "言語情報を用いた F_0 推定精度の向上," 日本音響学会平成4年度春季研究発表会講演論文集, 1-2-10, pp. 225-226 (1992.03).
 N. Kaiki, Y. Sagisaka: "An Improvement in F_0 Estimation Using Linguistic Information," ASJ Spring Meeting, 1-2-10, pp. 225-226 (1992.03).

Kaiki92ASJ03b 156

- 海木 延佳, 匂坂 芳典: “ポーズ長制御における個人性の分析,” 日本音響学会平成4年度春季研究発表会講演論文集, 1-2-23, pp. 251-252 (1992.03).
N. Kaiki, Y. Sagisaka: “Analysis of Personality for Pause Duration Control,” ASJ Spring Meeting, 1-2-23, pp. 251-252 (1992.03).
- Kaiki92SP03.....158
海木 延佳, 匂坂 芳典: “局所的句構造に基づくポーズ長の分析,” 電子情報通信学会技術研究報告, SP91-130, pp. 63-69 (1992.03).
N. Kaiki, Y. Sagisaka: “Analysis of Pause Duration Based on Local Phrase Structure,” IEICE Technical Report, SP91-130, pp. 63-69 (1992.03).
- Kaiki92IEICEtrans03.....165
海木 延佳, 武田 一哉, 匂坂 芳典: “言語情報を利用した母音継続時間長の制御,” 電子情報通信学会論文誌 A J75-A No. 3 pp. 467-473 (1992-03).
N. Kaiki, K. Takeda, Y. Sagisaka: “Vowel Duration Control Using Linguistic Information,” Trans. of IEICE, Vol. J75-A, No. 3, pp. 467-473 (1992.03).
- Kaiki92SP05.....172
海木 延佳, 匂坂 芳典: “局所的句構造に基づく F_0 制御,” 電子情報通信学会技術研究報告, SP92-6, pp. 41-46 (1992.05).
N. Kaiki, Y. Sagisaka: “ F_0 Control Based on Local Phrase Structure,” IEICE Technical Report, SP92-6, pp. 41-46 (1992.05).
- Kaiki92ASJ10.....178
海木 延佳, 匂坂 芳典: “発話速度による文音声のポーズ長変化の分析,” 日本音響学会平成4年度秋季研究発表会講演論文集, 1-5-16, pp. 247-248 (1992.10).
N. Kaiki, Y. Sagisaka: “Analysis of the Effect of Speaking rate on Pause Length in Sentence Speech Data,” ASJ Fall Meeting, 1-5-16, pp. 247-248 (1992.10).
- Kaiki92Book1.....180
Nobuyoshi Kaiki, Yoshinori Sagisaka: “Linguistic Properties in the Control of Segmental Duration for Speech Synthesis,” Talking Machines: Theories, Models & Applications, Elsevier-North-Holland Pub. pp. 255-263 (1992).
- Kaiki92Book2.....189
Nobuyoshi Kaiki, Yoshinori Sagisaka: “The Control of Segmental Duration in Speech Synthesis using Statistical Methods,” Speech Perception, Production and Linguistic Structure, OHM Pub. Tokyo pp. 391-402

(1992.01).

- Katagishi92ASJ10.....201
片岸 一起, 嵯峨山 茂樹, 相川 清明: “音声特徴パラメータ時系列の線形フィルタリングを用いたHMM音声認識,” 日本音響学会平成4年度秋季研究発表会講演論文集, 3-1-3, pp. 145-146, (1992.10).
K. Katagishi, S. Sagayama, K. Aikawa: “Linear Filtering of a Feature Parameter Sequence for HMM Speech Recognition,” ASJ Fall Meeting, 3-1-3, pp. 145-146, (1992.10).
- Katagishi92SST12.....203
Kazuki Katagishi, Harald Singer, Kiyooki Aikawa Shigeki Sagayama: “Linear filtering of a feature vector sequence for speech recognition,” Proc. of Fourth Australian International Conference on Speech Science and Technology, pp. 112-117, (1992.12).
- Kato92IEICE02.....209
加藤 喜永, 杉山 雅英: “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資料, SPREC-91-2, pp. 47-48 (1992.02).
Y. Kato, M. Sugiyama: “Continuous speech recognition using neural networks with multiple input-output units,” Continuous Speech Recognition Symposium, SPREC-91-2, pp. 47-48 (1992.02)
- Kato92ASJ03.....211
加藤 喜永, 杉山 雅英: “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,” 日本音響学会平成4年度春研究発表会講演論文集, 3-1-1, pp. 71-72 (1992.03).
Y. Kato, M. Sugiyama: “Continuous Speech Recognition Using Neural Networks with Multiple Input-output Units,” ASJ Spring Meeting, 3-1-1, pp. 71-72 (1992.03).
- Kato92IEICE06.....213
加藤 喜永, 杉山 雅英: “ファジィパーティションモデルを用いた連続音声認識,” 電子情報通信学会技術研究報告, SP92-28, pp. 31-38 (1992.06).
Y. Kato, M. Sugiyama: “Fuzzy Partition Models for Continuous Speech Recognition,” IEICE Technical Report, SP92-28, pp. 31-38 (1992.06).

Kato92NNSP08	220
Yoshinaga Kato and Masahide Sugiyama, "Fuzzy Partition Models and their effects in continuous speech recognition," IEEE Workshop Neural Networks Signal Processing, Elsinore, pp. 111-120 (1992.08).	
Kato92ASJ10	230
加藤 喜永, 杉山 雅英: "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," 日本音響学会平成4年度秋研究発表会講演論文集, 2-Q-20, pp. 211-212 (1992.10).	
Y. Kato, M. Sugiyama: "Speaker-independent Features Extracted by a Neural Network," ASJ Fall Meeting, 2-Q-20, pp. 211-212 (1992.10).	
Kato92ASJ(E)11	232
Yoshinaga Kato, Masahide Sugiyama: "Fuzzy Partition Models and Their Incremental Training for Continuous Speech Recognition," The Journal of the Acoustical Society of Japan(E), Vol. 13, No. 6 (1992.11).	
◆	
Komori92IEICEtrans	239
小森 康弘, Alexander H. Waibel, 嵯峨山 茂樹: "ニューラルファジー学習法による音声認識の性能向," 電子情報通信学会論文誌, D-II, Vol. J75-D-II, No. 7, pp. 1101-1110 (1992.07).	
Y. Komori, A. H. Waibel, S. Sagayama: "A Neural Fuzzy Training Approach for Improving Speech Recognition," Trans. of IEICE, D-II, Vol. J75-D-II, No. 7, pp. 1101-1110 (1992.07).	
Komori92ICASSP03	249
Yasuhiro Komori: "Neural Fuzzy Training Approach for Continuous Speech Recognition Improvement," Proc. of 1992 International Conference on Acoustics, Speech And Signal Processing, Vol. 1, S1, pp. I-405-408 (1992.03).	
Komori92ICSLP10	253
Yasuhiro Komori: "Minimum Error Classification Training for HMM-based Keyword Spotting," Proc. of 1992 International Conference on Spoken Language Processing, Tu.sAM.1.2, pp. 9-12 (1992.10).	

Kosaka92ASJ03	257
小坂 哲夫, 嵯峨山 茂樹: "不特定話者を対象とした混合連続分布 HMM 音声認識における混合数の検討," 日本音響学会平成4年度春季研究発表会講演論文集, 2-Q-20, pp. 197-198 (1992.03).	
T. Kosaka and S. Sagayama: "An Investigation on the Number of Mixture Components of Continuous HMMs for Independent Speakers," ASJ Spring Meeting, 2-Q-20, pp. 197-198, (1992.03).	
Kosaka92SP09	259
小坂 哲夫, 鷹見 淳一, 嵯峨山 茂樹: "話者混合 SSS による不特定話者音声認識と話者適応," 電子情報通信学会技術研究報告, SP92-52, pp. 17-24 (1992.09).	
T. Kosaka, J. Takami and S. Sagayama: "Speaker-Independent Speech Recognition and Speaker Adaptation Using Speaker-Mixture SSS algorithm," IEICE Technical Report, SP92-52, pp. 17-24, (1992.09).	
Kosaka92ASJ10a	267
小坂 哲夫, 嵯峨山 茂樹: "混合連続分布 HMM 音素モデルの構造自動決定法の検討," 日本音響学会平成4年度秋季研究発表会講演論文集, 2-1-1, pp. 79-80 (1992.10).	
T. Kosaka and S. Sagayama: "An Investigation on the Structure Determination of Continuous HMM Phone Models," ASJ Fall Meeting, 2-1-1, pp. 79-80 (1992.10).	
Kosaka92ASJ10b	269
小坂 哲夫, 鷹見 淳一, 嵯峨山 茂樹: "話者混合 SSS による不特定話者音声認識," 日本音響学会平成4年度秋季研究発表会講演論文集, 2-5-9, pp. 135-136 (1992.10).	
T. Kosaka, J. Takami and S. Sagayama: "Speaker-Independent Speech Recognition Using Speaker-Mixture SSS algorithm," ASJ Fall Meeting, 2-5-9, pp. 135-136, (1992.10).	
Kosaka92SST12	271
Tetsuo Kosaka, Shigeki Sagayama: "An Algorithm for Automatic HMM Structure Generation in Speech Recognition," Proc. of Fourth Australian International Conference on Speech Science and Technology pp. 104-109, (1992.12).	
◆	
KurinamiJASJ9211	277

K. Kurinami, M. Sugiyama: "An Optimization Technique for Speaker Mapping Neural Networks Using Minimal Classification Error Criterion," Journal Acoustic Society Japan, Vol. 13, No. 6, pp. 419-428 (1992.11).

K. Mimura, N. Kaiki, Y. Sagisaka: "An Evaluation of Clearness in the Synthetic Speech of ATR ν -Talk," ASJ Fall Meeting, 1-P-13, pp. 325-326 (1992.10).

◆
Lucke92ICASSP03.....286

Helmut Lucke, F. Fallside: "Expanding the Vocabulary of a Connectionist Recognizer Trained on the DARPA Resource Management Corpus," Proc. of 1992 International Conference on Acoustics, Speech, and Signal Processing, Vol. 1, S1, pp. I-605-608, San Francisco, (1992.03).

Lucke92ASJ10.....290

Helmut Lucke: "音声認識アルゴリズムに比較検討するための統計的考察," 日本音響学会平成4年度秋季研究発表会講演論文集, 2-Q-15, pp. 201-202 (1992.10).

H. Lucke: "Statistical Considerations in the Comparative Evaluation of Speech Recognition Algorithms," ASJ Fall Meeting, 2-Q-15, pp. 201-202 (1992.10).

Lucke92SP12.....292

Helmut Lucke: "Bayesian Causal Trees の理論を用いる確率文脈自由文法を推定方法," 電子情報通信学会技術研究報告, SP92-113, pp. 79-86 (1992.12)
H. Lucke: "A Method of Inferring Stochastic Context-free Grammars Using the Theory of Bayesian Causal Trees," IEICE Technical Report, SP92-113, pp. 79-86 (1992.12)

◆
Mimura92ASJ03.....300

三村 克彦, 海木 延佳, 匂坂 芳典: "統計的手法に基づくパワー制御規則の評価," 日本音響学会平成4年度春季研究発表会講演論文集, 1-2-15, pp. 235-236 (1992.03).
K. Mimura, N. Kaiki, Y. Sagisaka: "An Evaluation of Statistically Derived Rules for Amplitude Control" ASJ Spring Meeting, 1-2-15, pp. 235-236 (1992.03).

Mimura92ASJ10.....302

三村 克彦, 海木 延佳, 匂坂 芳典: "ATR ν -Talk 合成音声の明瞭度評価," 日本音響学会平成4年度秋季研究発表会講演論文集, 1-P-13, pp. 325-326 (1992.10).

◆
Miyazawa92ASJ10a.....304

宮沢 康永, 嵯峨山 茂樹: "移動ベクトル場平滑化話者適応方式における標準話者選択方式の検討," 日本音響学会平成4年度秋季研究発表会講演論文集, 2-5-2, pp. 121-122 (1992.10).

Y. Miyazawa and S. Sagayama: "An Examination for a Method of Standard Speaker Selection for Speaker Adaptation Based on Transfer Vector Field Smoothing Model," ASJ Fall Meeting, 2-5-2, pp. 121-122 (1992.10).

Miyazawa92ASJ10b.....306

宮沢 康永, 嵯峨山 茂樹: "全音素エルゴディックHMMを用いた教師なし話者適応方式," 日本音響学会平成4年度秋季研究発表会講演論文集, 2-5-3, pp. 123-124 (1992.10).

Y. Miyazawa and S. Sagayama: "Unsupervised Speaker Adaptation Using Ergodic HMM Consisting of All Phonemes in Speech Recognition," ASJ Fall Meeting, 2-5-3, pp. 123-124 (1992.10).

Miyazawa92SP10.....308

宮沢 康永, 大倉 計美, 嵯峨山 茂樹: "全音素エルゴディック HMM による教師なし話者適応," 電子情報通信学会技術研究報告, SP92-75, pp. 15-20 (1992.10).
Y. Miyazawa, K. Ohkura and S. Sagayama: "Unsupervised Speaker Adaptation Using All-phoneme Ergodic HMM," IEICE Technical Report, SP92-75, pp. 15-20, (1992.10).

Miyazawa92SST12.....314

Yasunaga Miyazawa, Shigeki Sagayama: "Speaker-normalized HMM-likelihood for Selecting a Reference Speaker in Speaker-adaptive Speech Recognition," Proc. of Fourth Australian International Conference on Speech Science and Technology pp. 431-436, (1992.12).

◆
Murakami92IEICetrans01.....320

村上 仁一, 荒木 哲郎, 池原 悟: "日本文音節入力に対して2重マルコフ連鎖モデルを用いた漢字かな交じり文節候補の抽出精度," 電子情報通信学会論文誌, D-II, Vol.

J75-D-II, No. 1, pp. 11-20 (1992.01).
 J. Murakami, T. Araki, S. Ikehara: "The Effects of Trigram Model in Japanese Speech Recognition," Trans. of IEICE, D-II, Vol. J75-D-II, No. 1, pp. 11-20 (1992.01).

Murakami92SPREC02..... 330
 村上 仁一, 嵯峨山 茂樹: "自由発話の音声認識の問題点について," 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資料, SPREC91-2, pp. 51 (1992.02).
 J. Murakami, S. Sagayama: ", " Continuous Speech Recognition Symposium, SPREC91-2, pp. 51 (1992.02).

Murakami92ASJ03..... 331
 村上 仁一, 嵯峨山 茂樹: "連続音声認識における LR パーザと単語 bigram の比較" 日本音響学会平成 4 年度春季研究発表会講演論文集, 1-P-8, pp. 119-120 (1992.03).
 J. Murakami, S. Sagayama: "Comparison Between LR Parser and Bigram Model for Speech Recognition," ASJ Spring Meeting, 1-P-8, pp. 119-120 (1992.10).

Murakami92SP10..... 333
 村上 仁一, 杉山 雅英, 渡辺 秀行: "HMM を用いた未知, 複数信号クラスタリング問題の検討," 電子情報通信学会信学技報, SP92-74, pp. 7-14 (1992.10).
 J. Murakami, M. Sugiyama, H. Watanabe: "Study of Unknown-Multiple Signal Source Clustering Problem Using Ergodic HMM," IEICE Technical Report, SP92-74, pp. 7-14 (1992.10).

Murakami92IPS10..... 341
 村上 仁一, 嵯峨山 茂樹: "HMM を用いた形態素解析," 情報処理学会第 45 回全国大会, pp. 3-161-162 (1992.10).
 J. Murakami, S. Sagayama: "Hidden Markov Model Applied to Morphological Analysis," pp. 3-161-162 (1992.10).

Murakami92ASJ10a..... 343
 村上 仁一, 嵯峨山 茂樹: "単語連鎖可変長統計の自動学習に基づく連続音声認識," 日本音響学会平成 4 年度秋季研究発表会講演論文集, 2-1-9, pp. 95-96 (1992.10).
 J. Murakami, S. Sagayama: "An Efficient Algorithm for Continuous Speech Recognition Using the Word Trigram Model," ASJ Fall Meeting, 2-1-9, pp. 95-96 (1992.10).

Murakami92ASJ10b..... 345

村上 仁一, 嵯峨山 茂樹: "単語の trigram を用いた連続音声認識のアルゴリズム," 日本音響学会平成 4 年度秋季研究発表会講演論文集, 2-Q-7, pp. 185-186 (1992.10).
 J. Murakami, S. Sagayama: "An Efficient Algorithm for Continuous Speech Recognition Using the Word Trigram Model," ASJ Fall Meeting, 2-Q-7, pp. 185-186 (1992.10).

Murakami92SST12..... 347
 Jin'ichi Murakami and Shigeki Sagayama: "An Efficient Algorithm For Using Word Trigram Models For Continuous Speech Recognition," Proc. of Fourth Australian International Conference on Speech Science and Technology, pp. 330-335 (1992.12).

◆

Nagai92ASJ03..... 353
 永井 明人, 菊池 英明, 嵯峨山 茂樹, 北 研二: "文脈自由文法から音素コンテキスト依存文法への変換アルゴリズム," 日本音響学会平成 4 年度春季研究発表会講演論文集, 3-1-6, pp. 81-82 (1992.03).
 A. Nagai, H. Kikuchi, S. Sagayama, K. Kita: "An Algorithm of Converting CFG into Phoneme-Context-Dependent Grammar," ASJ Spring Meeting, 3-1-6, pp. 81-82 (1992.03).

Nagai92SP06..... 355
 永井 明人, 鷹見 淳一, 嵯峨山 茂樹: "逐次状態分割法 (SSS) と音素コンテキスト依存 LR パーザを統合した SSS-LR 連続音声認識システム," 電子情報通信学会技術研究報告, SP92-33, pp. 69-76 (1992.06).
 A. Nagai, J. Takami, S. Sagayama: "The SSS-LR Continuous Speech Recognition System: Integrating SSS-derived Allophone Models and a Phoneme-Context-Dependent LR Parser," IEICE Technical Report, SP92-33, pp. 69-76 (1992.06).

Nagai92ICSLP10a..... 363
 Akito Nagai, Junichi Takami, Shigeki Sagayama: "The SSS-LR Continuous Speech Recognition System: Integrating SSS-derived Allophone Models and a Phoneme-Context-Dependent LR Parser," Proc. of 1992 International Conference on Spoken Language Processing, Vol. 2, pp. 1511-1514, Canada (1992.10).

Nagai92ICSLP10b..... 367

Akito Nagai, Kenji Kita, Toshiyuki Hanazawa, Tadashi Suzuki, Tomohiro Iwasaki, Takeshi Kawabata, Kunio Nakajima, Kiyohiro Shikano, Tsuyoshi Morimoto, Shigeki Sagayama, Akira Kurematsu: "Hardware Implementation of Realtime 1000-word HMM-LR Continuous Speech Recognition," Proc. 1992 International Conference on Spoken Language Processing, Vol. 1, pp. 237-240, Canada (1992.10).

Nagai92ASJ10.....371

永井 明人, 鷹見 淳一, 嵯峨山 茂樹: "逐次状態分割法 (SSS) と LR パーザを統合した SSS-LR 連続音声認識手法の文音声による評価," 日本音響学会平成4年度秋季研究発表会講演論文集, 2-1-8, pp. 93-94 (1992.10).

A. Nagai, J. Takami, S. Sagayama: "Evaluation of Continuous Uttered Sentence Recognition by SSS-LR Integrated Successive State Splitting Algorithm with an LR Parser," ASJ Fall Meeting, 2-1-8, pp. 93-94 (1992.10).



NakamuraICASSP9203.....373

Satoru Nakamura, Hidefumi Sawai, Masahide Sugiyama: "Speaker-Independent Phoneme Recognition Using Large-Scale Neural Networks," Proc. of 1992 International Conference on Acoustics, Speech And Signal Processing, 46.5, pp. I-409-412 (1992.03).



Ohkura92SPREC02.....377

大倉 計美, 杉山 雅英, 嵯峨山 茂樹: "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資料, SPREC-91-2, pp. 49-50 (1992.02).

K. Ohkura, M. Sugiyama, S. Sagayama: "Speaker Adaptation Based on Moving Vector Field Model with Continuous Mixture Density HMMs," Continuous Speech Recognition Symposium, pp. 49-50 (1992.02).

Ohkura92ASJ03a.....379

大倉 計美, 土井 啓輔, 杉山 雅英: "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," 日本音響学会平成4年度春季研究発表会講演論文集, 2-Q-16, pp. 189-190, (1992.03).

K. Ohkura, K. Doi, M. Sugiyama: "Noisy Speech Recognition Based

on Transfer Vector Field Smoothing Adaptation," ASJ Spring Meeting, 2-Q-16, pp. 189-190 (1992.03).

Ohkura92ASJ03b.....381

大倉 計美, 杉山 雅英, 嵯峨山 茂樹: "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," 日本音響学会平成4年度春季研究発表会講演論文集, 2-Q-17, pp. 191-192 (1992.03).

K. Ohkura, M. Sugiyama, S. Sagayama: "Speaker Adaptation Based on Transfer Vector Field Smoothing Model with Continuous Mixture Density HMMs," ASJ Spring Meeting, 2-Q-17, pp. 191-192, (1992.03).

Ohkura92SP06.....383

大倉 計美, 杉山 雅英, 嵯峨山 茂樹: "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," 電子情報通信学会技術研究報告, SP92-16, pp. 23-28, (1992.06).

K. Ohkura, M. Sugiyama, S. Sagayama: "Speaker Adaptation Based on Transfer Vector Field Smoothing Model with Continuous Mixture Density HMMs," IEICE Technical Report, SP92-16, pp. 23-28, (1992.06).

Ohkura92ICSLP10.....389

Kazumi Ohkura, Masahide Sugiyama, Shigeki Sagayama: "Speaker Adaptation Based on Transfer Vector Field Smoothing with Continuous Mixture Density HMMs," Proc. of 1992 International Conference on Spoken Language Processing, Vol. 1, pp. 369-372 (1992.10).

Ohkura92ASJ10a.....393

大倉 計美, David Rainton, 杉山 雅英: "識別誤り規準を用いた耐雑音 HMM の検討," 日本音響学会平成4年度秋季研究発表会講演論文集, 1-7-14, pp. 73-74 (1992.10).

K. Ohkura, D. Rainton, M. Sugiyama: "Noise-robust HMMs Based on Minimum Error Classification," ASJ Fall Meeting, 1-7-14, pp. 73-74, (1992.10).

Ohkura92ASJ10b.....395

大倉 計美, 杉山 雅英, 嵯峨山 茂樹: "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," 日本音響学会平成4年度秋季研究発表会講演論文集, 2-5-4, pp. 125-126 (1992.10).

K. Ohkura, M. Sugiyama, S. Sagayama: "Evaluation of Speaker Adaptation Method Based on Transfer Vector Field Smoothing Model with Continuous Mixture Density HMMs," ASJ Fall Meeting, 2-5-4, pp. 125-126, (1992.10).

Ohkura92IEEEWorkshop10	397
Kazumi Ohkura, Masahide Sugiyama: "Noisy Speech Recognition with Codebook Mapping and Segmental Approaches," Proc. of IEEE Workshop on Interactive Voice Technology for Telecommunications Applications, VII.3, New Jersey (1992.10).	
◆	
Rainton92SP01	402
David Rainton, Shigeki Sagayama: "Minimum Error Classification Training of HMMs - Implementation Details and Experimental Results," 電子情報通信学会技術研究報告, SP91-107, pp. 39-46 (1992.01). IEICE Technical Report	
Rainton92ASJ03	410
David Rainton, Shigeki Sagayama: "Word Level Minimum Error Training of Phoneme HMMs," 日本音響学会平成4年度春季研究発表会講演論文集, 1-1-2, pp. 3-4 (1992.03). ASJ Spring Meeting	
Rainton92ISSPA08	412
David Rainton, Shigeki Sagayama: "A New Minimum Error Classification Training Technique for HMM Based Speech Recognition," Proc. of ISSPA92, pp. 279-282, Australia (1992.08)	
Rainton92ASJ10	416
David Rainton, Shigeki Sagayama: "Sentence Level Minimum Error Training of Phoneme HMMs," 日本音響学会平成4年度秋季研究発表会講演論文集, 2-Q-13, pp. 197-198 (1992.10). ASJ Fall Meeting	
Rainton92ICSLP10	418
David Rainton, Shigeki Sagayama: "Appropriate Error Criterion Selection For Continuous Speech HMM Minimum Error Training," Proc. of 1992 International Conference on Spoken Language Processing, We.fAM.1.3, pp. 233-236 (1992.10).	
Rainton92JASJ11	422

David Rainton, Shigeki Sagayama: "Minimum Error Classification Trainign of HMMs- Implementation details and experimental results," The Journal of the Acoustical Society of Japan, Vol. 13, No. 6, pp. 379-387 (1992.11).



Sagayama92JASJ01	431
嵯峨山 茂樹: "数理統計モデルを用いた音声認識の現状と将来," 日本音響学会誌, 48巻, 1号, pp. 26-32, (1992.01). S. Sagayama: "A Discussion on Statistical Approaches to Speech Recognition," The Journal of the Acoustical Society of Japan, Vol. 48, No. 1, pp. 26-32 (1992.01).	
Sagayama92SpeechTech02	438
Shigeki Sagayama: "Speech Research at ATR Interpreting Telephony Research Laboratories," Speech Technology, Vol. 5, No. 4, pp. 23-27, Media Dimension Publishing Co., Ltd., (1992.02).	
Sagayama92Sympo02	443
嵯峨山 茂樹: "音声認識における韻律情報の利用について," 連続音声認識シンポジウム (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」時限研究会主催, 於浜松館山寺) 予稿集 (1992.02). S. Sagayama: "On the Use of Prosodic Information," Proc. of the Symposium on Continuous Speech Recognition (Kanzanji)(1992.02).	
Sagayama92kakenhi2	446
嵯峨山 茂樹: "音声認識のための確率モデル, ニューラルネットワーク, 話者適応, 韻律情報の利用等に関する研究," 文部省科研費「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」報告書 (1992.02). S. Sagayama: "Researches of Probabilistic Models, Neural Networks, Speaker Adaptation, and Prosody for Speech Recognition," <i>The Final Report of New Speech Recognition Paradigms Including Markov Models and Neural Networks</i> , Science Research Fund from the Ministry of Education and Culture (1992.02).	
Sagayama92NikkeiByte06	459
嵯峨山茂樹: "音声認識," 日経バイト6月号 (100号記念特集号), 日経マグロービル社, pp. 212-221 (1992.06). S. Sagayama: "Speech Recognition," Proc. of Nikkei Byte, pp. 212-221 (1992.06).	

Sagayama92SST12.....	469
Shigeki Sagayama, Masahide Sugiyama, Kazumi Ohkura, Jun'ichi Takami, Akito Nagai, Harald Singer, Hiroaki Hattori, Keiji Fukuzawa, Yasunaga Kato, Kouichi Yamaguchi, Jin'ichi Murakami, Akira Kurematsu: "ATREUS: Continuous Speech Recognition Systems at ATR Interpreting Telephony Research Laboratories," Proc. of SST92 (Fourth Australian International Conference on Speech Science and Technology) (Brisbane), pp. 324-329 (1992.12).	
◆	
Sagisaka92ICASSP03.....	475
Yoshinori Sagisaka, Nobuyoshi Kaiki: "Optimization of Intonation Control Using Statistical F_0 Resetting Characteristics," Proc. of 1992 International Conference on Acoustics, Speech, and Signal Processing, Vol.II, pp. 49-52 (1992.03).	
Sagisaka92ICSLP10.....	479
Yoshinori Sagisaka, Naoto Iwahashi, Katsuhiko Mimura: "ATR ν -TALK Speech Synthesis System," Proc. of 1992 International Conference on Spoken Language Processing, Th.fAM.2.2, pp. 483-486 (1992.10).	
Sagisaka92SPREC10.....	483
匂坂 芳典, 海木 延佳, 岩橋 直人, 三村 克彦: "ATR ν -TALK 音声合成システム," 電子情報通信学会「音声認識の実用化を目指す新手法」時限研究専門委員会資料, SPREC-92-2 (1992.10).	
Y. Sagisaka, N. Kaiki, N. Iwahashi, K. Mimura: "ATR ν -TALK Speech Synthesis System," SPREC-92-2 (1992.10).	
Sagisaka92JASJ12.....	487
匂坂 芳典, 浦谷則好: "ATR 音声・言語データベース," 日本音響学会誌, 48 巻, 12 号, pp. 878-882 (1992.12).	
Y. Sagisaka, N. Uratani: "ATR Spoken Language Database," The Journal of the Acoustical Society of Japan, Vol. 48, 12, pp. 878-882 (1992.12).	
◆	
Singer92ICASSP03.....	492

Harald Singer, S. Sagayama: "Pitch Dependent Phone Modelling for HMM Based Speech Recognition," Proc. of 1992 International Conference on Acoustics, Speech, and Signal Processing, 36.1, pp. 273-276,(San Francisco) (1992.03).	
Singer92ASJ03.....	496
Harald Singer, 嵯峨山 茂樹: "Matrix Parsing Applied to TDNN-Based Speech Recognition," 日本音響学会平成 4 年度春季研究発表会講演論文集, 3-1-10, pp. 89-90 (1992.03).	
H. Singer, S. Sagayama: "Matrix Parsing Applied to TDNN-Based Speech Recognition," ASJ Spring Meeting, 3-1-10, pp. 89-90 (1992.03).	
Singer92SP10.....	498
Harald Singer, 嵯峨山 茂樹: "行列パーザとその HMM 音声認識への応用," 電子情報通信学会技術研究報告, SP92-76, pp. 21-26, (1992.10).	
H. Singer, S. Sagayama: "Matrix Parser and its Application to HMM-based Speech Recognition," IEICE Technical Report, SP92-76, pp. 21-26, (1992.10).	
Singer92ASJ10a.....	504
Harald Singer, 嵯峨山 茂樹: "行列パーザとその HMM 音声認識への応用," 日本音響学会平成 4 年度秋季研究発表会講演論文集, 2-1-3, pp. 83-84 (1992.10).	
H. Singer, S. Sagayama: "Matrix Parsing Applied to HMM-Based Speech Recognition," Proc. ASJ Fall Meeting, 2-1-3, pp. 83-84 (1992.10).	
Singer92ASJ10b.....	506
シンガー ハラルド, 相川清明: "時間周波数マスキング特性を用いた混合連続 HMM による連続音声認識," 日本音響学会平成 4 年度秋季研究発表会講演論文集, 1-1-2, pp. 3-4, (1992.10).	
Harald Singer, Kiyooki Aikawa: "Time-Frequency Masking Characteristics applied to Continuous Speech Recognition with Continuous Mixture HMM's," Proc. ASJ Fall Meeting, 1-1-2, pp. 3-4, (1992.10).	
Singer92SST12.....	508
Harald Singer, Shigeki Sagayama: "Suprasegmental Duration Control with Matrix Parsing in Continuous Speech Recognition," Fourth Australian International Conference on Speech Science and Technology (Brisbane), pp. 394-399, (1992.12).	

Sugiyama92LECO2.....	514
杉山 雅英: “ニューラルネットワークの音声認識への応用,” システム総合研究所シンポジウム, ニューラルネットワークの新展開と応用, S92-2, pp. 67-123 (1992.02).	
Sugiyama92SP02.....	516
杉山 雅英: “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).	
M. Sugiyama: “Several Properties of LPC Cepstrum Vector Space,” IEICE Technical Report, SP91-120, pp. 17-24 (1992.02).	
Sugiyama92ASJ03a.....	524
杉山 雅英, 栗並 賢太郎: “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,” 日本音響学会平成 4 年度春季研究発表会講演論文集, 2-Q-22, pp. 201-202 (1992.03).	
M. Sugiyama, K. Kurinami: “An Optimization of Speaker Mapping Neural Network Using Mimimal Classification Error Criterion,” ASJ Spring Meeting, 2-Q-22, pp. 201-202 (1992.03).	
Sugiyama92ASJ03b.....	526
杉山 雅英, 吉田 哲也: “スペクトル遷移の確率モデルによる言語識別の検討,” 日本音響学会平成 4 年度春季研究発表会講演論文集, 1-P-1, pp. 105-106 (1992.03).	
M. Sugiyama, T. Yoshida: “Language Identification Using Probabilistic Spectral Transition Modelling,” ASJ Spring Meeting, 1-P-1, pp. 105-106 (1992.03).	
Sugiyama92SRS06.....	528
Masahide Sugiyama: “LANGUAGE RECOGNITION USING SPECTRAL FEATURES,” Proc. of Speech Research Symposium, Rutgers University (1992.06).	
Sugiyama92NNSP08.....	532
Masahide Sugiyama, Kentaro Kurinami: “Minimal Classification Error Optimization for a Speaker Mapping Neural Network,” Proc. of NNSP92, pp. 233-242 (1992.8).	
Sugiyama92IEICE09.....	542
杉山 雅英, 渡辺 秀行, 村上 仁一: “事前登録なしの話者識別問題,” 電子情報通信学会秋季大会, SA-6-1, 1-263-264 (1992.09).	

M. Sugiyama, H. Watanabe, J. Murakami: “Speech Classification for Speaker-Specific Difference Without Pre-Registration,” SA-6-1, 1-263-264 (1992.09).	
Sugiyama92ASJ10a.....	544
杉山 雅英, 村上 仁一, 渡辺 秀行: “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,” 日本音響学会平成 4 年度秋季講演論文集, 2-Q-21, pp. 213-214 (1992.10).	
M. Sugiyama, J. Murakami, H. Watanabe: “Speech Segmentation and Clustering Based on Speaker Features,” ASJ Fall Meeting, 2-Q-21, pp. 213-214 (1992.10).	
Sugiyama92ASJ10b.....	546
杉山 雅英, 大倉 計美: “レベル可変雑音モデルによる雑音抑圧法,” 日本音響学会平成 4 年度秋季研究発表会講演論文集, 1-7-12, pp. 69-70 (1992.10).	
M. Sugiyama, K. Ohkura: “Level Variable Noise Model for Noise Restoration,” ASJ Fall Meeting, 1-7-12, pp. 69-70 (1992.10).	
◆	
Takami92ICASSP03.....	548
Jun-ichi Takami, Shigeki Sagayama: “A Successive State Splitting Algorithm for Efficient Allophone Modeling,” Proc. of 1992 International Conference on Acoustics, Speech, and Signal Processings, 66.6, San Francisco (1992.03).	
Takami92ASJ3a.....	552
鷹見 淳一, 嵯峨山 茂樹: “逐次状態分割法 (SSS) により自動生成した隠れマルコフ網の性能評価,” 日本音響学会平成 4 年度春季研究発表会講演論文集, 2-1-8, pp. 65-66, (1992.03).	
J. Takami, S. Sagayama: “Evaluation of the Hidden Markov Network Generated Automatically by the Successive State Splitting Algorithm,” ASJ Spring Meeting, 2-1-8, pp. 65-66, (1992.03).	
Takami92ASJ3b.....	554
鷹見 淳一, 嵯峨山 茂樹: “隠れマルコフ網 (HM-Net) を用いた話者適応,” 日本音響学会平成 4 年度春季研究発表会講演論文集, 1-1-8, pp. 15-16, (1992.03).	
J. Takami, S. Sagayama: “Speaker Adaptation Using the Hidden Markov Network,” ASJ Spring Meeting, 1-1-8, pp. 15-16, (1992.03).	
Takami92SP9.....	556

- 鷹見 淳一: “音声認識における HMM とその精度向上のための手法,” 電子情報通信学会技術研究報告, SP92-49, pp. 17-24 (1992.9).
- J. Takami: “The Hidden Markov Model on Speech Recognition and Appro aches to Improve the Model Accuracy,” Technical Report of IEICE, SP92-49, (1992.09).
- Takami92ASJ10a.....564
- 鷹見 淳一, 小坂 哲夫, 嵯峨山 茂樹: “話者方向を加えた逐次状態分割法 (SSS) による話者共通隠れマルコフ網の生成,” 日本音響学会平成 4 年度秋季研究発表会講演論文集, 3-1-8, pp. 155-156(1992.10).
- J. Takami, T. Kosaka, S. Sagayama: “Automatic Generation of Speaker-common Hidden Markov Network by Adding the Speaker Splitting Domain to the Successive State Splitting Algorithm,” ASJ Fall Meeting, 3-1-8, pp. 155-156, (1992.10).
- Takami92ASJ10b 566
- 鷹見 淳一, 永井 明人, 嵯峨山 茂樹: “逐次状態分割法 (SSS) と LR パーザを統合した SSS-LR 連続音声認識手法における話者適応の性能評価,” 日本音響学会平成 4 年度秋季研究発表会講演論文集, 2-5-5, pp. 127-128 (1992.10).
- J. Takami, A. Nagai, S. Sagayama: “Evaluation of Speaker Adaptation on the SSS-LR Continuous Speech Recognition Technique Integrating the Successive State Splitting Algorithm with the LR Parser,” ASJ Fall Meeting, 2-5-5, pp. 127-128 (1992.10).
- Takami92JASJ11.....568
- Jun-ichi Takami, A. Kai, S. Sagayama: “A Pairwise Discriminant Approach Using Artificial Neural Networks for Continuous Speech Recognition,” The Journal of the Acoustical Society of Japan, Vol. 13, No. 6, pp. 411-418, (1992.11).
- Takami92JATR..... 576
- 鷹見 淳一, 永井, 嵯峨山 茂樹: “aka(赤) と aki(秋) の /k/ は同じ音? - 前後の音素を考慮した高精度音声認識 -,” ATR ジャーナル, No. 12 (1992.11).
- J. Takami, A. Nagai, S. Sagayama: “Is the /K/ in Aka the same sounds the /K/ in Aki?-Speech Recognition Using Phoneme Context Information,” ATR Journal, No. 12, (1992.11).
- Takami92SST12..... 582
- Jun-ichi Takami, Akito Nagai, Shigeki Sagayama: “Speaker Adaptation of the SSS (Successive State Splitting)-Based Hidden Markov Network for Continuous Speech Recognition,” Proc. of SST92, pp. 437-442, Australia (1992.12).
- Tamato92SIG04.....588
- 田本 真詞, 村上 仁一, 嵯峨山 茂樹: “HMM を利用した言語獲得の可能性について,” 人工知能学会研究会資料, SIG-SLUD-9201-6, pp. 47-54, (1992.04).
- S. Tamoto, J. Murakami, S. Sagayama: “Investigation of Language Model Using HMM,” Special Interesting Group on Spoken Language Understanding of Japanese Society of AI, SIG-SLUD-9201-6, pp. 47-54 (1992.04).
- Wang92ASJ10.....596
- 王文俊, Nick. Campbell, 岩橋 直人, 匂坂 芳典: “Unit Selection for English Speech Synthesis Using Regression Trees,” 日本音響学会平成 4 年度秋季研究発表会講演論文集, 1-5-10, pp. 235-236, (1992.10).
- Wern Jun Wang, W. N. Campbell, Naoto Iwahashi, Yoshinori Sagisaka, ASJ Fall Meeting, 1-5-10, pp. 235-236, (1992.10).
- Watanabe92SP07 598
- 渡辺 秀行, 村上 仁一, 杉山 雅英: “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 —,” 電子情報通信学会技術研究報告, SP92-45, pp. 47-54 (1992.07).
- H. Watanabe, J. Murakami, M. Sugiyama: “Unknown-Multiple Signal Source Clustering Problem —An Application to Unknown Speaker Clustering —,” IEICE Technical Report, SP92-45, pp. 47-54 (1992.07).
- Willems92SP12 606
- ウィレムス エドワード, 小坂 哲夫, 鷹見 淳一, 嵯峨山 茂樹: “音声認識のための隠れマルコフ網の動的話者適応法,” 信学技報, SP92-102, pp. 57-64, (Dec. 1992).
- E. Willems, T. Kosaka, J. Takami and S. Sagayama: “A Dynamic Approach to Speaker Adaptation of Hidden Markov Networks for Speech Recognition,” IECIE Technical Report, SP92-102, pp. 57-64, (Dec. 1992).

- ◆
- Yamaguchi92ASJ03a.....614
 山口耕市, 嵯峨山茂樹: “混合連続分布型 HMM を用いた HMM-LR 連続音声認識,” 日本音響学会平成 4 年度春季研究発表会講演論文集, 1-P-5, pp. 113-114, (1992.03).
 K. Yamaguchi, S. Sagayama: “HMM-LR Continuous Speech Recognition Using Continuous Mixture HMMs,” ASJ Spring Meeting, 1-P-5, pp. 113-114 (1992.03).
- Yamaguchi92ASJ03b 616
 山口耕市, 嵯峨山茂樹, 北研二, Frank K. Soong: “HMM-LR 連続音声認識における A* アルゴリズムを用いた探索手法の検討,” 日本音響学会平成 4 年度春季研究発表会講演論文集, 3-1-9, pp. 87-88, (1992.03).
 K. Yamaguchi, S. Sagayama, K. Kita, Frank K. Soong: “A Study on the Search Method Using A* Algorithm in HMM-LR Continuous Speech Recognition,” ASJ Spring Meeting, 3-1-9, pp. 87-88, (1992.03).
- Yamaguchi92ICSLP10 618
 Kouichi Yamaguchi, Shigeki Sagayama, Kenji Kita, Frank K. Soong: “Continuous Mixture HMM-LR Using the A* Algorithm for Continuous Speech Recognition,” Proc. of 1992 International Conference on Spoken Language Processing, We.sAM.1.2, pp. 301-304, (1992.10).
- Yamaguchi92ASJ10a.....622
 山口耕市, 嵯峨山茂樹: “ニューラルネットワークを用いた適応的ビームサーチによる HMM-LR 連続音声認識,” 日本音響学会平成 4 年度秋季研究発表会講演論文集, 3-7-2, pp. 165-166 (1992.10).
 K. Yamaguchi, S. Sagayama: “A Neural Network Controlled Adaptive Beam Search Method for HMM-LR Continuous Speech Recognition,” ASJ Fall Meeting, 3-7-2, pp.165-166, (1992.10).
- Yamaguchi92ASJ10b 624
 山口耕市, 永井明人, 鷹見淳一, 大倉計美, 小坂哲夫, 福沢圭二, 加藤喜永, 北研二, Harald Singer, 村上仁一, 杉山雅英, 嵯峨山茂樹, 服部浩明, 小森康弘, 沢井秀文, 花沢利行, 中村哲, 甲斐充彦, 南泰浩, 川端豪, 鹿野清宏: “ATREUS: ATR における連続音声認識諸方式の比較,” 日本音響学会平成 4 年度秋季研究発表会講演論文集, 2-Q-5, pp. 181-182 (1992.10).
 K. Yamaguchi, A. Nagai, J. Takami, K. Ohkura, T. Kosaka, K. Fukuzawa, Y. Kato, K. Kita, H. Singer, J. Murakami, M. Sugiyama, S. Sagayama,

H. Hattori, Y. Komori, H. Sawai, T. Hanazawa, S. Nakamura, A. Kai, Y. Minami, T. Kawabata, K. Shikano: “Comparisons between Continuous Speech Recognition Systems (ATREUS) at ATR,” ASJ Fall Meeting, 2-Q-5, pp. 181-182, (1992.10).

◆

- Yamamoto92IPS10.....626
 山本 寛樹, 村上 仁一, 嵯峨山 茂樹: “隠れマルコフモデルによる言語モデルの自動獲得の検討,” 情報処理学会第 45 回全国大会講演論文集, pp. 3-227-228 (1992.10).
- H. Yamamoto, J. Murakami, S. Sagayama: “A Preliminary Study on Language Acquisition Using Hidden Markov Model,” Proc. of the 45th National Convention IPS Japan, 6G-6, pp. 3-227-228 (1992.10).
- Yamamoto92ASJ10 628
 山本 寛樹, 村上 仁一, 嵯峨山 茂樹: “HMM を言語モデルに用いた連続音声認識の検討,” 日本音響学会平成 4 年度秋研究発表会講演論文集, 2-Q-11, pp. 193-194 (1992-10).
 H. Yamamoto, J. Murakami, S. Sagayama: “A Study on Continuous Speech Recognition Using HMM Language Model,” ASJ Fall Meeting, 2-Q-11, pp. 193-194 (1992.10).

1993年 発表論文リスト

ATR 自動翻訳電話研究所 音声情報処理研究室

以下のリストは、ATR 自動翻訳電話研究所 音声情報処理研究室が、1993年1月から3月の間に発表、出版、あるいは投稿されだあるいはされる予定の論文や学会発表のリストである。一部の学会に関しては、概要を投稿した段階であるものもある。現時点で収集できた情報であるので、抜けている情報があるかも知れない点を了承願いたい。

- [Campbell93Phonetica] W. N. Campbell: "Prosodic Encoding of English Speech," Phonetica (submitted)
- [Donescu93IEICE01] ドネスク ヨアナ, 加藤 喜永, 杉山 雅英, "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評価," 信学技法, SP92-117, pp. 15-22 (1993-01).
- [Fujiwara93ASJ3] 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," 音学講論, 2-Q-12, pp. ???-??? (1993.3).
- [Fujiwara93IEICE1] 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," 信学技報, SP92-132, pp.73-80 (1993.1).
- [Fukuzawa93ASJ03] 福沢 圭二, 杉山 雅英: "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," 日本音響学会平成5年度春季研究発表会講演論文集, 2-Q-20, (1993.03).
- [Fukuzawa93IEICE-DII] 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, D-II, 投稿中 (1993年2月 原稿提出).
- [Hattori93IEICEjournal] H. Hattori and S. Sagayama: "Speaker Adaptation Based on Vector Field Smoothing," IEICE Trans Inf. & Syst., Vol. E76-D, No. 2., Feb 1993.
- [Hattori93IEICEjournal] H. Hattori, S. Nakamura, K. Shikano, and S. Sagayama: "Speaker Weighted Training of HMM Using Multiple Reference Speakers," 電子通信学会英文誌, Vol.E, Feb 1993.
- [Iwashita93ASJ3] 岩橋 直人, 匂坂 芳典: "空間多重分割型数量化法による英語音声のセグメント継続時間制御モデル," 日本音響学会平成5年度春季研究発表会講演論文集, 3-8-4, (1993-3).
- [Katagishi93SC] Kazuki KATAGISHI, Harald SINGER, Kiyooki AIKAWA Shigeki SAGAYAMA: "Linear filtering of a feature vector sequence for speech recognition," submitted to Speech Communication.
- [Kita93IEICEjournal] K. Kita, T. Morimoto, and S. Sagayama: "LR Parsing with a Category Reachability Test Applied to Speech Recognition," IEICE Trans Inf. &

Syst., Vol. E76-D, No. 1, pp. 23-28, Jan 1993.

- [Kosaka93ASJ03] 小坂 哲夫, エドワード・ウイレムス, 鷹見 淳一, 嵯峨山 茂樹, "複数の話者適応に基づく動的な話者適応," 音講論集, 掲載予定.
- [Kosaka93Euruspeech] Tetsuo Kosaka, Edward Willems, Jun-Ichi Takami, Shigeki Sagayama: "A Dynamic Approach to Speaker Adaptation of Hidden markov Networks for Speech Recognition," Proc. of Eurospeech93, proposal 投稿中.
- [Kosaka93ICASSP04] T. Kosaka, S. Sagayama: "Rapid Speaker Adaptation Using Speaker-Mixture Allophone Models Applied for Speaker Independent Recognition," Proc. ICASSP93.
- [Kosaka93ICASSP] Tetsuo Kosaka, Jun-Ichi Takami, Shigeki Sagayama: "Rapid Speaker Adaptation Using Speaker-Mixture Allophone Models Applied to Speaker-Independent Speech Recognition," Proc. of ICASSP93, 掲載予定.
- [Lucke93EUROSP09] H.Lucke: "Inference of stochastic context-free grammar rules from example data using the theory of Bayesian Belief propagation," Eurospeech (Berlin), in preparation, (1993.09)
- [Lucke93JOURNAL] H.Lucke: "An Application of Bayesian belief Propagation in causal trees to the inference of stochastic context-free grammar rules from example text," In preparation, (1993)
- [Mimura93ASJ3] 三村, 岩橋, 匂坂: "単位接続歪が合成音声明瞭度に与える影響について," 音学講論 (1993.03)(予定).
- [Miyazawa93ASJ03] 宮沢 康永, 嵯峨山 茂樹: "全音素エルゴディック HMM を用いた教師なし話者適応における初期モデルの検討," 日本音響学会平成5年度春季研究発表会講演論文集, 2-4-10, pp. 37-38 (1993.03).
- [Miyazawa93FULLPAPER] 宮沢 康永, 大倉 計美, 嵯峨山 茂樹: "全音素エルゴディック HMM を用いた教師なし話者適応," 電子情報通信学会論文誌 D-II, 投稿中.
- [Miyazawa93ICASSP04] Y. Miyazawa, (S. Sagayama): "An All-Phoneme Ergodic HMM for Unsupervised Speaker Adaptation," Proc. ICASSP93.
- [Muarakami93Eurospeech] Jin'ichi Murakami and Hiroaki Yamamoto and Shigeki Sagayama: "On the Automatic Acquisition of Stochastic Network Grammar using Ergodic HMM," Eurospeech93 proposal (1993).
- [Murakami93SIG02] 村上, 山本, 嵯峨山: "Ergodic HMM による確率つきネットワーク文法の獲得の可能性について," 人工知能学会, SIG-SLUD-9204-3 (1993-3).
- [Nagai93FULLPAPERb] 永井 明人, 北 研二, 花沢 利行, 鈴木 志, 岩崎知弘, 川端 豪, 中島 邦男, 鹿野 清宏, 森元 暹, 嵯峨山 茂樹, 樽松 明: "HMM と一般化 LR 構文解析を用いた長時間大語彙連続音声認識装置の実現," 日本音響学会 投稿予定
- [Nagai93FULLPAPERc] 永井, 鷹見, 嵯峨山: "隠れマルコフ網と一般化 LR 構文解析を統合した連続音声認識," 電子情報通信学会 投稿予定

- [Nagai93ICASSP] A. Nagai, K. Yamaguchi and A. Kurematsu, "ATREUS : A Comparative Study of Continuous Speech Recognition Systems at ATR," Proc. ICASSP93 (Minneapolis, USA), (1993.4 発表予定).
- [Nagai93IEICEjournal] A. Nagai, S. Sagayama, K. Kita, and H. Kikuchi: "Three Different LR Parsing Algorithms for Phoneme-Context-Dependent HMM-Based Continuous Speech Recognition," IEICE Trans Inf. & Syst., Vol. E76-D, No. 1, pp. 29-37, Jan 1993.
- [Nagai93IEICetrans01] Akito Nagai, Shigeki Sagayama, Kenji Kita, Hideaki Kikuchi: "Three Different LR Parsing Algorithms for Phoneme-Context-Dependent HMM-based Continuous Speech Recognition", Trans. of IEICE, (電子情報通信学会論文誌 音声対話処理特集号)(1993.01).
- [Nagai93SP1] 永井明人, 山口耕市, 鷹見淳一, 大倉計美, 小坂哲夫, 福沢圭一, 加藤喜永, Harald Singer, 村上仁一, 杉山雅英, 嵯峨山茂樹, 保坂順子, 森元暹, 北研二(徳島大学), 服部浩明(日本電気), 小森康弘(キヤノン), 沢井秀文(リコー), 花沢利行(三菱電機), 中村哲(シャープ), 甲斐充彦(豊橋技科大), 南泰浩, 川端豪, 鹿野清宏(以上NTT), 樽松明, "ATRにおける連続音声認識システム'ATREUS'の諸方式と性能," 信学技報, SP92-122, pp. 51-58, (1993.1).
- [Sagayama93ASJ3] 嵯峨山, 鷹見, 永井, H. Singer, 谷戸, 鈴木, 森元, 樽松, "自動翻訳電話実験システム ASURA の概要," 音学講論 (1993.3 出版予定).
- [Sagayama93Eurospeech] S. Sagayama, J. Takami, A. Nagai, H. Singer, K. Yamaguchi, K. Ohkura, K. Kita, and A. Kurematsu, "ATREUS: a Speech Recognition Front-end for a Speech Translation System," Proc. Eurospeech93 (Berlin), (proposal submitted).
- [Sagayama93FULLPAPER] S. Sagayama, A. Nagai, J. Takami, K. Yamaguchi, H. Singer, M. Sugiyama, H. Hattori(+), K. Ohkura, K. Fukuzawa, Y. Kato, T. Kosaka, J. Murakami, and A. Kurematsu: "ATREUS: continuous speech recognition systems at ATR Interpreting Telephony Research Laboratories," submitted to Speech Communication
- [Sagayama93book] 嵯峨山 茂樹、鹿野 清宏、他: "音声認識," 「翻訳電話」(樽松 明編), ATR 先端技術シリーズ, オーム社, 1993(?).
- [Sagayama93ISCIE03] 嵯峨山 茂樹: "最近の音声認識技術と自動翻訳電話," システム制御情報学会 研究会資料, Mar 1993.
- [Sagayama93Eurospeech] S. Sagayama, J. Takami, A. Nagai, H. Singer, K. Yamaguchi, K. Ohkura, K. Kita, and A. Kurematsu, "ATREUS: a Speech Recognition Front-end for a Speech Translation System," Proc. Eurospeech93 (Berlin), (proposal submitted).
- [Singer93ICASSP03] H.Singer, S.Sagayama: "Matrix Parser and its Application to HMM-based Speech Recognition," ICASSP93 (Minneapolis), pp. ???-???, (1993.03)
- (to appear).
- [Takami93ASJ03] 鷹見 淳一, 宮沢 康永, 永井 明人, 嵯峨山 茂樹, "話者適応型 SSS-LR 連続音声認識方式における標準話者予備選択の効果," 音講論集, 2-Q-16 (1993.03).
- [Takami93FULLPAPER] 鷹見 淳一, 嵯峨山 茂樹: "逐次状態分割法による隠れマルコフ網の自動生成," 電子情報通信学会 投稿予定