

TR-I-0005

連続音声認識

Continuous Speech Recognition

鹿野 清宏

Kiyohiro Shikano

1987.6

あらまし

本報告は、連続音声認識へのアプローチの一例として、

- (1) スペクトログラム・リーディングに基づく音韻認識、
- (2) ブラックボード・モデルに基づくシステム構成、
- (3) ベクトル量子化を介した話者適応、
- (4) 単語Trigram Model による統計的言語処理、

を取り上げて、著者が関連してきた研究を主体として、平易に解説したものである。上記の分野に関連する文献も多く示すようにして、連続音声認識の研究の最近の一面を把握するのに役立つことを目指した。

ATR 自動翻訳電話研究所

(ATR Interpreting Telephony Research Laboratories)

連続音声認識

鹿野 清宏
(ATR自動翻訳電話研究所)

1. まえがき

人は、日常生活で、自分の意志を音声で伝えているので、究極的には、計算機等の機械とのコミュニケーションも音声で行えることが望ましい。人間は、頭の中で考えた事柄や感じたことを、概念から単語列に変換し、さらに、発声器官を駆動して声に出して、他人に伝える。聞き手は、会話状況、文脈、意味、構文、単語等の言語情報を駆使して、話し手の音声を理解し、意図を把握している。

このように、音声認識の過程を、人間の言語活動の一環として総合的に捕らえようという音声理解の研究が、1970年台の前半に、米国のDARPA(Defence Advanced Research Projects Agency)の第一次音声理解システムプロジェクトで取り上げられた⁽⁴⁾。1976年に、CMU(Carnegie-Mellon University)の連続音声認識システムHarpyが一応DARPAの定めた1000単語からなる連続音声の対象の目標を達成した⁽¹³⁾。また、同じCMUで研究されたHearsay IIは、ブラックボードモデルと呼ばれる新しいAIシステム構築の概念を提案したものの、計算機設備の不十分さと音声知識の不完全さのために成功には至らなかった⁽¹⁰⁾。

日本でも、ほぼ同時期に、NTT 武蔵野通研⁽¹¹⁾⁽¹⁷⁾⁽²²⁾⁽²³⁾、京都大⁽¹⁶⁾、山梨大⁽²¹⁾、京都工繊大⁽¹⁸⁾で連続音声認識システムの作成が試みられた。NTT 武蔵野通研のシステムは、新幹線の座席予約を対象として、予約項目ごとに区切って発声した音声を認識し、会話モデルに従って合成音声で応答するものであった。1980年に入ると、音声認識の研究の重点は、LSI技術の発展に伴い、マッチングベースの単語音声認識装置の開発に重点が置かれた。しかしながら、研究者の数も増加し、特徴ベースの音韻認識、話者の適応化手法、不特定話者音声認識、大語彙単語音声認識等の基礎的研究も続けられ、技術の蓄積が行われた。

1985年から、米国では、DARPAのStrategic Computing Projectの一つとして、CMU, BBN(Bolt Beranek and Newman Inc.), MIT(Massachusetts Institute of Technology)を中心に第2次の音声理解系の研究が開始されている⁽¹⁾⁽⁵⁾⁽³⁶⁾。そのプロジェクトでは、言語処理技術、特徴ベースの認識手法、分散システム構築技術を融合して、音声理解システムを構築を目指すとともに、音韻ラベルのつけられた音声データベースの蓄積にも重点を置いている。ヨーロッパでも、英国のアルペイ計画、EC諸国のエスプリ計画の新世代計算機プロジェクトが活発に行われており、そのプロジェクトの一部に音声認識/理解の研究も含まれている。

我が国では、昭和61年度から、国際電気通信基礎技術研究所(ATR International)のATR自動翻訳電話研究所で、音声理解を含む自動翻訳電話のプロジェクトが開始された。ここでは、自動翻訳電話システムの構築のための、音声認識、機械翻訳、音声合成、システム構築技術、音声データベース、言語データベース等の要素技術の基礎研究が開始されている⁽³⁶⁾。表1-1に音声理解システムに関連する主な研究機関およびプロジェクトを示しておく。

連続音声中の音韻は、次にあげる様々な変形を伴っている。

- (1) 音韻環境の影響(音韻環境による調音結合、異音等)。
- (2) 発声環境、マイクロフォン等の影響。
- (3) 発声速度、発声方法等、話者内での発声の違い。
- (4) 発声者による違い(声道長、発声法、方言等)。

表 1-1 音声理解に関する内外の研究動向

機関 あるいは プロジェクト	国名	研究目標
ATR自動翻訳電話研究所	日 本	自動翻訳電話システムの要素技術の基礎研究
DARPA (CMU, MIT, BBN, TI, SRI, etc.)	米 国	不特定大語彙連続音声認識システム
IBM	米 国	オフィス文書の大語彙音声入力システム(単 語音声認識ベース)
ベル研究所	米 国	単語音声認識システムの基礎研究
アルペイ計画	英 国	次世代コンピュータの研究と関連して音声認 識、音声合成の高性能化
エスプリ計画	EC諸国	情報通信技術と関連して音声認識、音声合成 の高性能化

これらの項目について研究が進められているけれども、全く基礎的研究レベルである。本稿では、主に、(1)の音韻認識と(4)の不特定話者音声認識の問題に対する一アプローチについて、例をあげて説明する。また、外国語を聞くときに経験するように、言語知識なしに音韻を聞き取ることは、きわめて困難である。この言語処理との関連した問題への一アプローチとしての統計的な言語処理モデルについても簡単な例を示す。

2. 音韻認識

精度よく音韻を認識するには、調音結合、異音等の音韻知識を十分に考慮する必要がある。図2-1は、「*We were at work all week.*」と発声したときの音声スペクトログラムである。この音声には、*/w/*, */l/*, */r/* の半母音と、2つの */k/* 及び、単語境界でしばしば現れる */t/* の異音が含まれている。このような音声スペクトログラムに対しては、音韻の場所を同定し音韻識別を行うという単純な過程では、音韻の認識を行うことは、非常に困難である。視察によるスペクトログラムリーディングは、次のように行われる⁽¹⁵⁾。

/w/ や */l/* は、第1ホルマント(F1)と第2ホルマント(F2)が1kHz以下の低い周波数から後続母音にゆっくりと移行する特徴として見つけ出される。また、*/r/* は、F3が2kHz以下に下がってゆくという動きの特徴として見つけ出すことができる。さらに、*/t/* は、単語境界での */t/* の異音として、声帯の動きの停止として存在の可能性が推測される。2つの */k/* は、閉塞区間に続くバーストのスペクトルから、2つのバーストのスペクトルがかなり異なっているものの、存在の可能性が推測される。さらに、前母音(front vowel) */i, e/* もF1とF2の間のギャップから、比較的容易に推測することができる。音韻認識システムを、少なくともこの様な音韻知識を利用して、構築する必要がある。そのためには、セグメンテーション → 音韻識別というボトムアップだけでなく、音韻マッチング → セグメンテーションのトップダウン過程や音形規則の活用も不可欠である。当然のことながら、音韻単位、CV単位という固定した単位に基づくスペクトルマッチングのみで音韻認識を行うことは無理と考えられる。

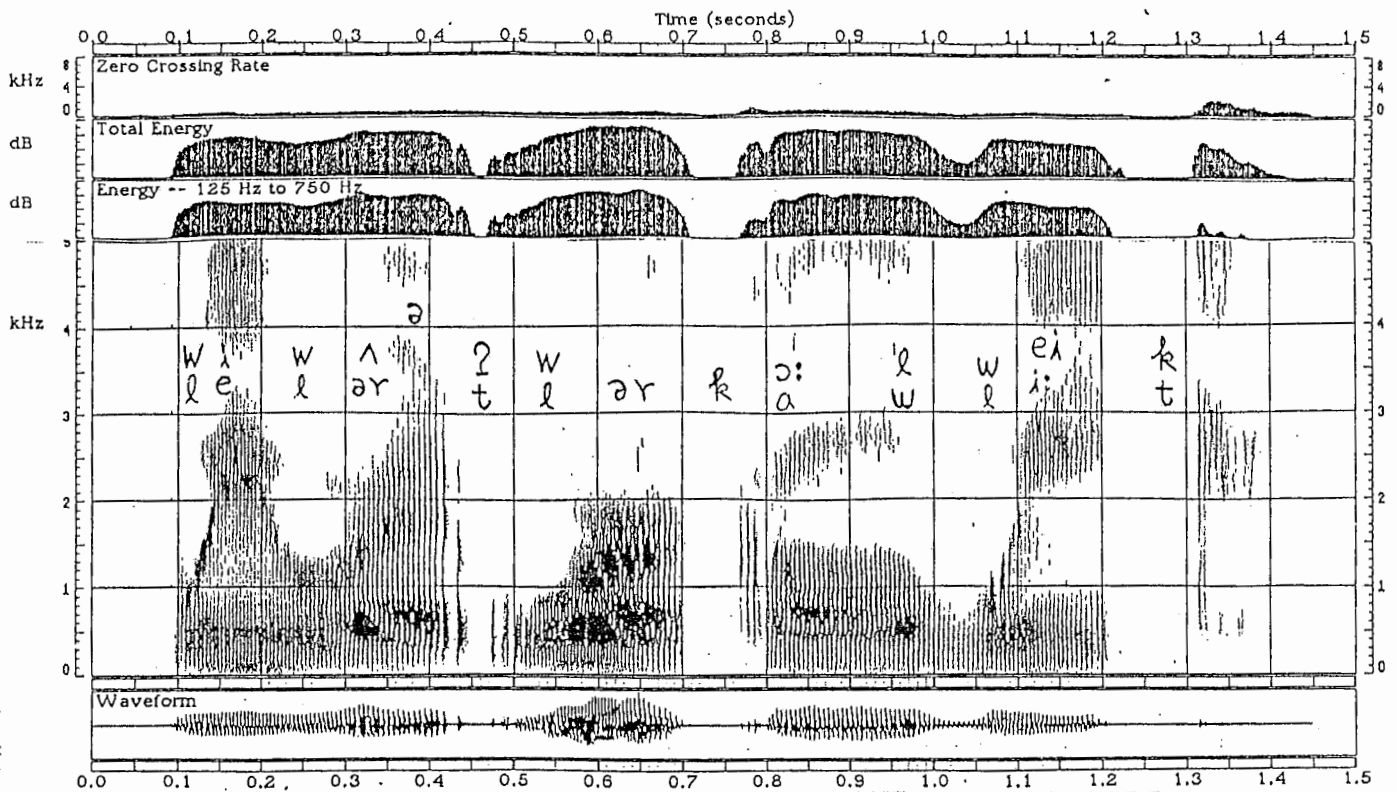


図2-1. 音声スペクトログラムと音韻認識の例(「We were at work all week.」)

このような考え方に基づく音韻認識への一アプローチとして、文献(14)で提案されているトップダウン処理とボトムアップ処理の融合をあげることができる。図2は、「arigatai」と発声した音声のスペクトログラムである。ボトムアップ処理によって、子音 /r/ /t/、そして、文頭の /a/ 等は認識されるが、/i ga/, /ai/ は、各々 VCV あるいは、VV の音韻連鎖パターンとして処理するのが適切である。このような音韻連鎖知識に基づく VCV や VV の処理は、音韻の組み合わせの数が多くなりがちであるので、上位の音韻辞書のからの仮説によるトップダウン的処理によって行うことが適切である。さらに、このような仮説検定の過程を記憶して、同じ環境での処理を繰り返さないようにすることにより、処理速度の向上を図ることも可能である。

このような過程で、音声スペクトルとの対応を検定する手法として、

- (1) 特徴ベースによる手法、
- (2) スペクトルマッチングによる手法、

がある。両方の手法を用いたシステムとして、文献(14)が上げられる。特徴ベースによる手法については、MIT の Zue 等によるスペクトログラムのリーディング特徴についての研究や⁽¹²⁾⁽¹⁵⁾⁽²⁶⁾⁽²⁷⁾、DARPA のプロジェクトでの CMU の Ron Cole のグループの研究⁽¹⁾⁽⁷⁾が参考になる。また、図2-1と図2-2の音声に対する音韻認識のアプローチから推測できるように、英語は、子音の特徴が比較的是っきりしている言語であり、一方、日本語は、母音の特徴が比較的明確な言語である。このような言語の違いにも留意して研究を進めることも重要である。

スペクトルマッチング尺度は、LPC分析に基づく、LPCケプストラム距離、最尤スペクトル距離から、スペクトルのピークに重みをおくことのできるWLR尺度⁽³⁷⁾、メルスケールのように周波数軸に重みを置くメルケプストラム距離、周波数軸重み付けWLR尺度⁽³⁸⁾と研究が進展してきた。また、パワー情報の利用⁽³⁴⁾や、差分スペクトル情報⁽³⁹⁾の有効性も示されている。

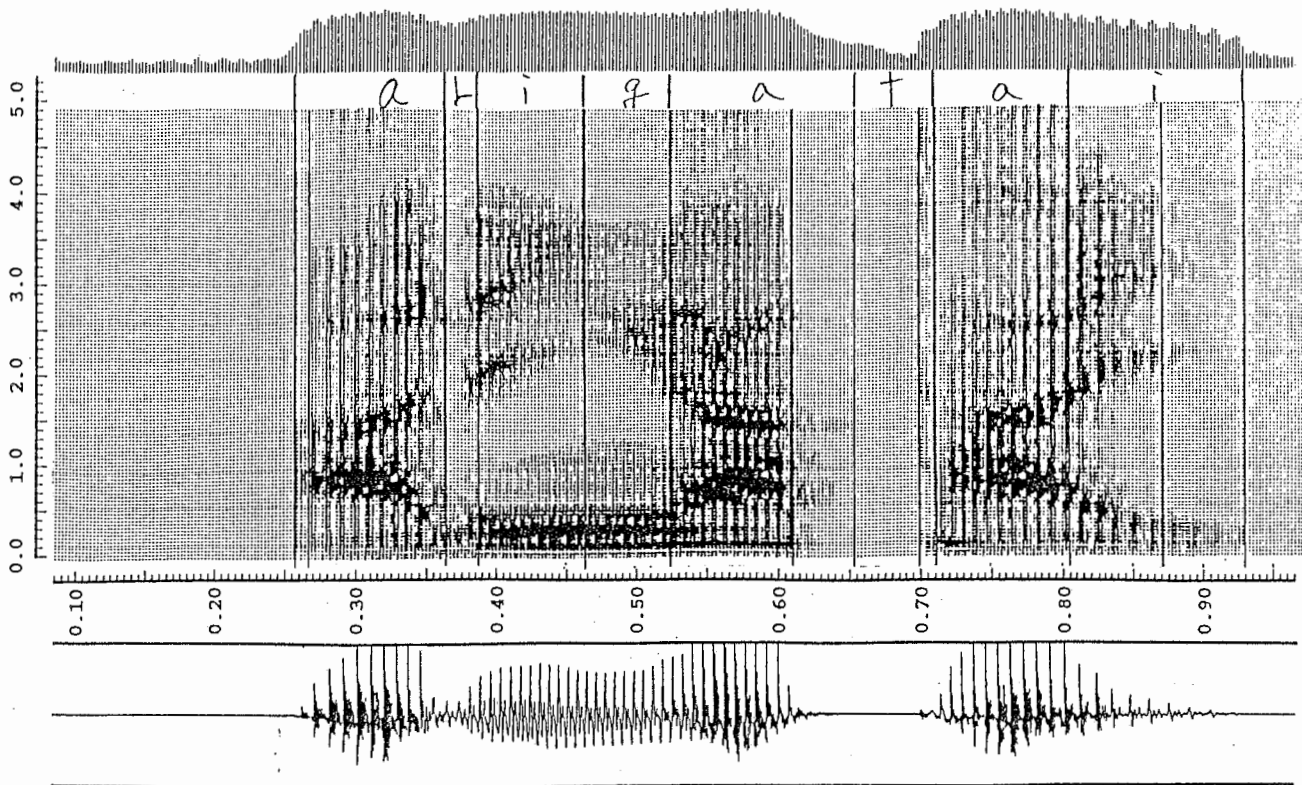


図2-2. 音声スペクトログラムと音韻認識の例(「arigatai」)

今後、スペクトル尺度の研究は、単語認識での距離尺度というようなグローバルな観点からよりも、各音韻ごとに詳しく尺度の有効性を検討すべきである⁽⁴⁰⁾。また、耐雑音性や聴覚の観点からの、よりターゲットを絞った研究も必要である。

3. 連続音声認識システム

連続音声認識システムを実現するには、音韻に関する知識の外に、音韻情報の曖昧さを補うために、図3-1にしめすように言語情報を最大限に利用する必要がある。我々が外国語を聞く場合、言語情報なしに音韻を書き留めることができないことを思い浮かべれば、言語情報の利用が連続音声認識システムの構築に不可欠であることが理解できよう。

図3-2は、現在、ATR自動翻訳電話研究所で研究を開始している連続音声認識システムの一構成例である。システムは、並列処理、データフロー処理が容易であるボトムアップ処理により、認識処理の核となる認識結果を生成し、その後、トップダウン処理により、より詳しい精密な認識処理を行ってゆく。制御方式は、オブジェクト指向的概念を取り入れ、各処理部を Actor を生成する KS (Knowledge Source) と考える。各 Actor は、各分散ブラックボードを共有することができるし、各分散ブラックボードを介して外の Actor や KS と情報の授受を行えるようにすることを考える。処理の高速化にも、各処理部 (KS) に置かれる分散ブラックボードを活用する。以下、各処理部について、簡単に説明を加える。

特徴抽出部

LPC分析情報をベースとし、スペクトログラム情報、波形情報を補助的に用いて、音韻認識のための特徴パラメータを計算する。その外、ピッチ周波数やストレスの強さ等の韻律的特徴も計算する。このモジュールは、音韻認識部、ワードスッポティング部等の上位の処理部からもトップダウン的に起動されることもできる。抽出された特徴パラメータは、分散ブラックボードの特徴パラメータに格納される。

音韻認識部

スペクトログラムリーディング情報に基づく音韻認識アルゴリズムと、LPC スペクトル情報を用いたパターンマッチングに基づく音韻認識アルゴリズムを組み合わせる。音韻認識は、特徴抽出部の特徴パラメータを、主に用いて、実行されるが、特徴抽出部モジュールへのトップダウン的起動も頻繁に行われる。音韻認識の結果は分散ブラックボードの音韻ラティスに格納される。

ワードスポッティング部

音韻ラティスを基にして、キーワードのワードスポッティングを行う。音韻ラティス情報が不明確である場合には、音韻認識モジュールを介して、特徴抽出部の特徴パラメータを参照したり、特徴抽出モジュールを起動したりする。ワードスポッティングの過程で、韻律情報の活用も図る。キーワードのワードスポッティングに基づいて、文節処理部を起動し、活用形の予測情報や後続単語の予測情報を受け取り、連続単語認識処理を実行する。ワードスポッティングの結果は、分散ブラックボードの単語ラティスに蓄えられる。文節の認識結果は、単語グラフとして、文節処理部の分散ブラックボードに蓄えられる。

文節処理部

ワードスポッティング部からの単語認識結果に基づき、構文、意味、分散ブラックボードの韻律情報を駆使して、後続の単語を予測して、その情報をワードスポッティング部に送り返す。構文、意味情報の利用に際して、単語連鎖⁽²⁵⁾や係り受け連鎖⁽¹⁹⁾の統計情報も用いる。動的計画法や、木探索法を用いて効率化を図ってゆく。

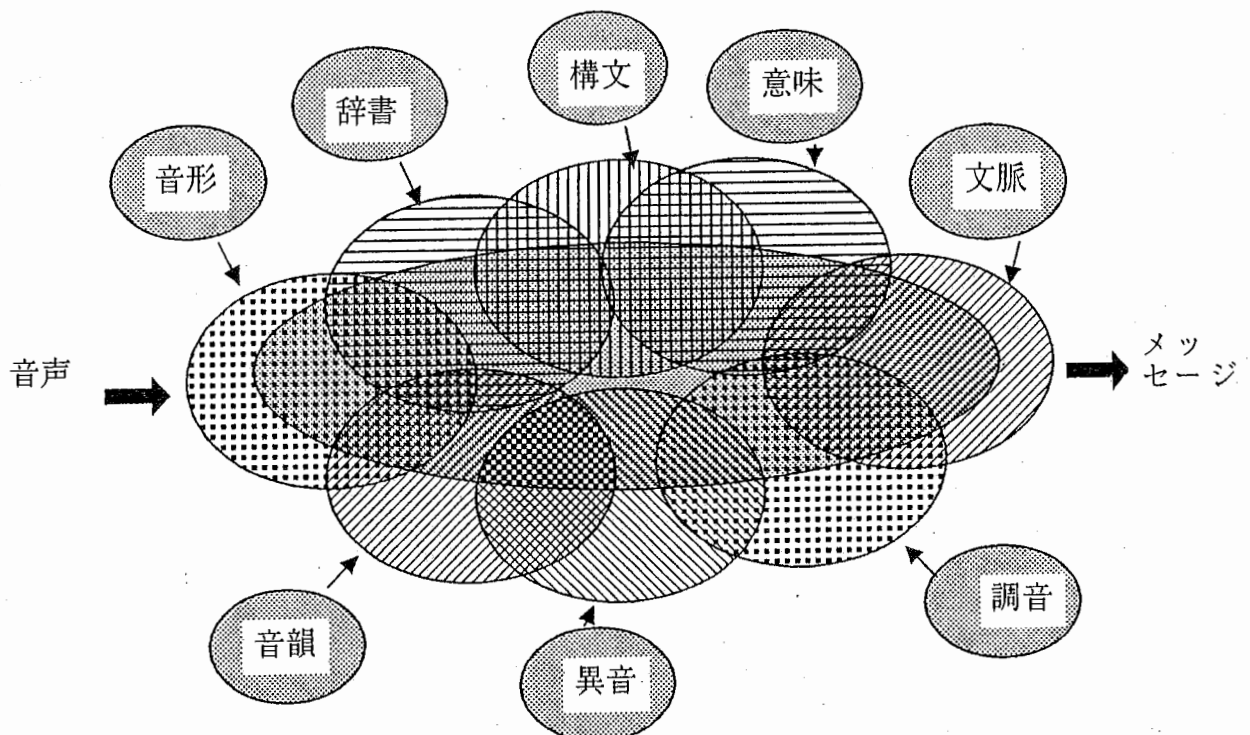


図3-1 連続音声認識

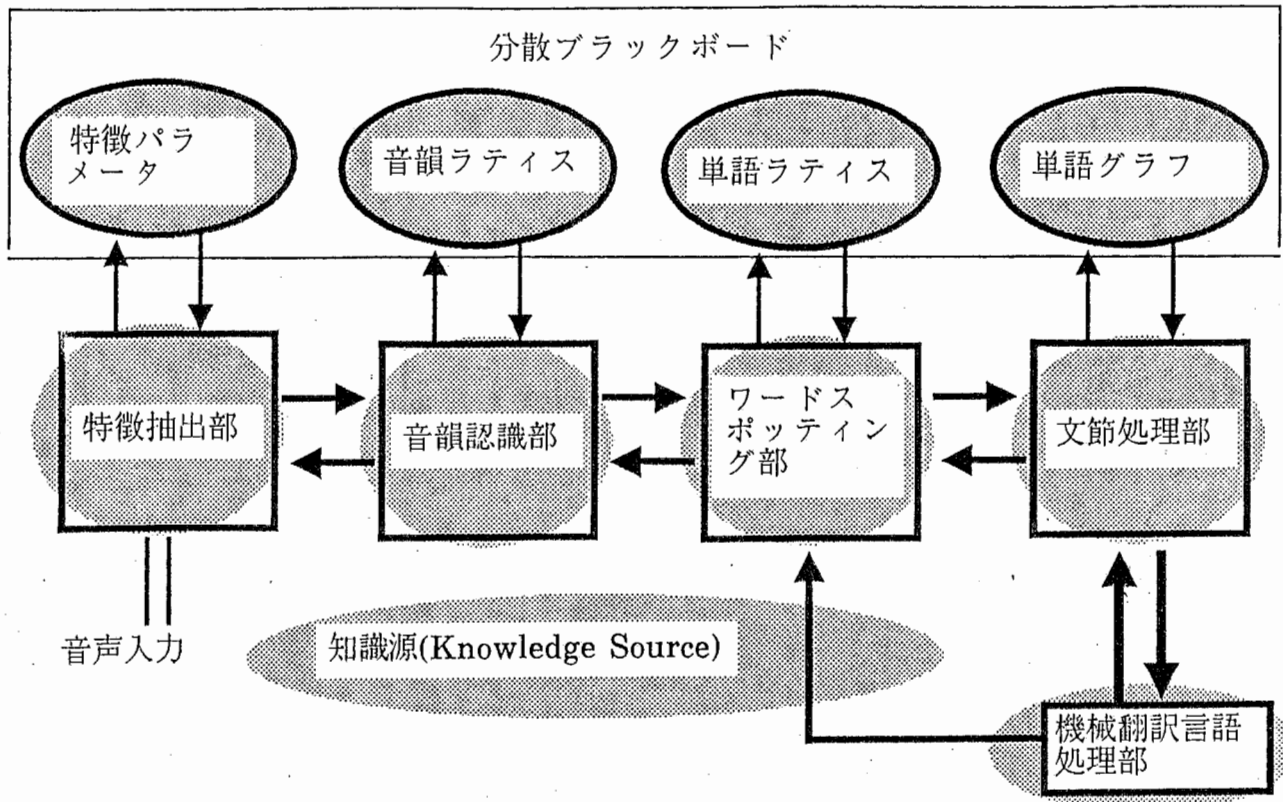


図3-2 音声認識処理システムの構成例

4. 話者適応

不特定話者大語彙音声認識への現実的なアプローチの一つとして、話者適応からのアプローチが有望と思われる。ここでは、このアプローチの一例として、ベクトル量子化を介した話者適応化の考え方について簡単に紹介する⁽²⁴⁾。

不特定話者音声認識の問題に対処するには、少なくとも次の2つの話者正規化の問題を考慮に入れる必要がある。

- (a) スペクトルの正規化。
- (b) 発声、調音の正規化。

(b)の発声や調音の違い等の話者適応化が困難である問題に対し、ここでは、単語単位のマルチテンプレートとして対処する。(a)の分析フレームごとのスペクトルの正規化の問題を、標準話者と発声者間のスペクトルの変換を求める問題と考える。両者間のスペクトルの変換は、複雑であり簡単な線形な変換で表すことができない。ベクトル量子化による手法は、両者のスペクトル変換を、2つの符号帳間のスペクトルの対応として表す。

ベクトル量子化は、LPC分析を介した音声スペクトルを、256個程度の少ない数のLPCスペクトルで近似する手法である。まず、標準話者と発声者のスペクトルを各々、256個程度のLPCスペクトルで近似する。その上で、図4-1に示すような発声者の符号帳のスペクトルと、標準話者の符号帳のスペクトル間の対応関係を見出す。この対応関係を求めるために、学習用に発声した単語音声間の対応をDPマッチングを介してヒストグラムを採り、スペクトル間の対応を求めたり、発声者から標準話者への(あるいは標準話者から発声者への)スペクトルの変換に用いる平均スペクトルを求める。

この手法により、単語ごとに区切って発声されたIBMのOffice Correspondence Task⁽²¹⁾ (422単語)を用いた単語音声認識実験で、48.1%の単語認識率が、上記の適応化によ

り、表4-1に示すように80.9%まで改善され、ベクトル量子化介した話者適応化アルゴリズムの有効性が示されている。図4-2に、このベクトル量子化介した話者適応化を用いたDouble SPLIT法⁽³³⁾による単語音声認識システムのブロック図を示しておく。

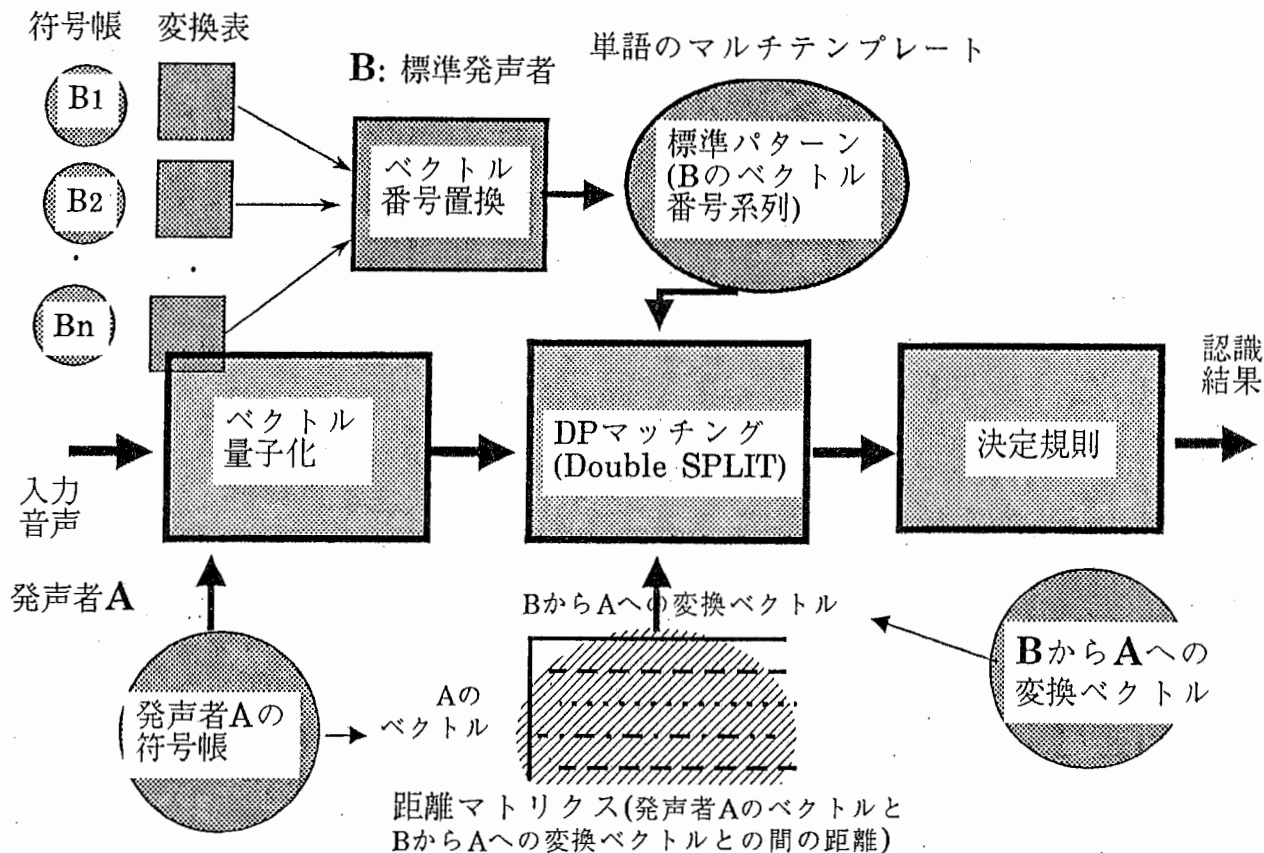


図4-2 ベクトル量子化を介した話者適応による単語音声認識システム

表4-1 ベクトル量子化を介した話者適応による単語音声認識実験結果

	単語音声認識実験結果		
	= 1	≤2	≤3
average of three	80.9%	92.0%	95.5%
speaker : AGC (American male)	78.4%	91.0%	95.9%
speaker : KEN (Australian male)	80.4%	89.8%	92.6%
speaker : BSF (American female)	84.2%	95.4%	98.0%

5. 統計的言語処理

言語は、言語固有のエントロピーを持っている。音声認識において認識誤りを訂正したり、次に続く単語列を予測するのに、このエントロピーの冗長性を利用することが有

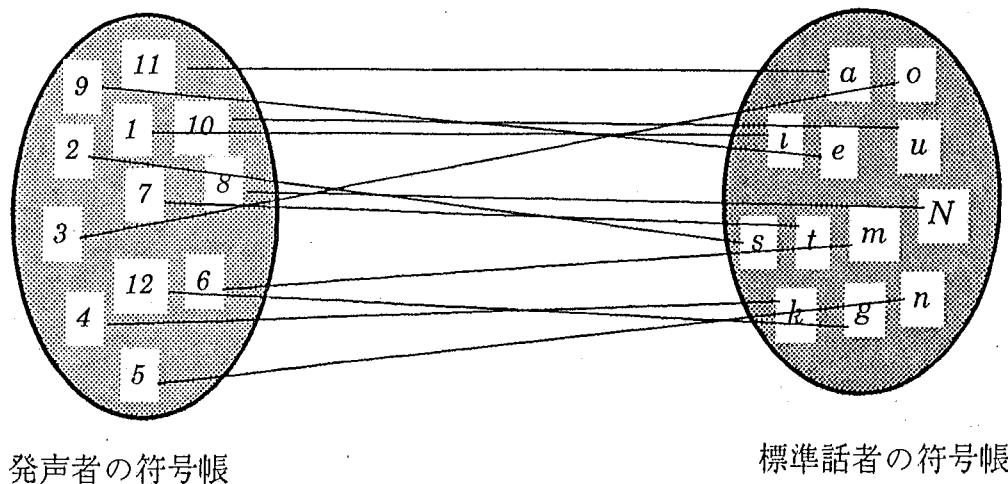


図 4-1 ベクトル量子化によるスペクトルの正規化

効と考えられる。ここでは、このようなアプローチの一例として、単語の3つ組の生起確率を利用した Trigram モデルによる単語音声認識結果の改善の例を紹介する⁽²⁶⁾。

認識対象は、前の章と同じ単語ごとに区切って発声されたIBMの Office Correspondence Task である。Trigram モデルへの入力は、表4-1の認識実験結果で、第10位までの単語と各々の尤度である。Trigramの確率の推定のために、英語の代表的なテキストデータベースである Brown Corpus⁽²⁹⁾を用いた。Brown Corpus は、約1,000,000 単語を含む文章に、約90余りの単語の品詞や属性を付けたテキストデータベースである。表5-1に、Trigram モデルによる単語認識率の向上の様子を話者ごとに示しておく。平均で、80.9%の単語認識率が88.5%まで向上し、この Trigram モデルの有効性がわかる。IBMのOffice Correspondence Taskは、話し言葉に近い文章であり、一方、Brown Corpus テキストデータベースは、1961年に出版された本のテキストからなるデータベースである。両者の文体には大きな違いがあり、もし、大量のIBMのタスクのテキストデータベースを用いることができれば、さらに、単語認識率が改善されることが期待される。表5-2に、誤り訂正の例を、単語認識での候補とともに示しておく。

6. むすび

連続音声認識への技術は、まだまだ、基礎的レベルにあり、多くの基礎的研究を必要としている。ここで、連続音声認識システムの研究の方向についての私見を述べ、むすびとする。

音声認識の研究レベルを効率よく向上させるために、音韻のラベル等が付けられた音声データベース作成し、データベース管理システム、音声研究用のワークベンチを開発する。多量の音声データベースを用い音韻認識、単語認識等の認識技術の評価を行う。さらに、個々の認識技術の適用範囲、問題点なども明確にする。

音韻認識技術の精密化、高度化のために、パターンマッチングのみならず聴覚、スペクトルリーディング等からの多面的なアプローチが必要である。

大語彙連続音声認識を行うには、音韻認識に基づくワードスポッティングを核にして認識システムを構築する。

むやみに大語彙不特定話者認識システムを目指すのではなく、話者適応化からアプローチする。

表5-1 Trigram モデルによる単語認識誤りの訂正結果

[発声者3人の平均]	単語認識率
Without Trigram	80.9%
With Trigram	
$P(w_3 w_1w_2)$	88.5%
$P(w_3 w_1w_2), P(w_2w_3 w_0w_1), P(w_2w_3w_4 w_0w_1)$	89.1%

[Speaker: AGC, American Male]	
Without Trigram	78.4%
With Trigram	
$P(w_3 w_1w_2)$	87.4%
$P(w_3 w_1w_2), P(w_2w_3 w_0w_1), P(w_2w_3w_4 w_0w_1)$	88.7%

[Speaker: KEN, Australian Male]	
Without Trigram	80.4%
With Trigram	
$P(w_3 w_1w_2)$	86.6%
$P(w_3 w_1w_2), P(w_2w_3 w_0w_1), P(w_2w_3w_4 w_0w_1)$	86.2%

[Speaker: BSF, American Female]	
Without Trigram	84.2%
With Trigram	
$P(w_3 w_1w_2)$	91.4%
$P(w_3 w_1w_2), P(w_2w_3 w_0w_1), P(w_2w_3w_4 w_0w_1)$	92.7%

音声を認識するには、意味、構文はもとより文脈、状況等の言語情報を最大限に利用する必要があり、認識対象を限定し、理解システムの立場から研究を進める。また、言語固有のエントロピーを利用した統計的言語処理の研究を進めるためにも大規模なテキストデータベースを作成する必要がある。

連続音声認識システムの実現には、様々な知識源からなるモジュールを密に結び付ける必要があり、システム構成技術が発展し、並列処理、分散処理のためのオペレーティングシステムやツールを整備する必要がある。

参考文献

- (1) D.A.Adams, R.Bisiani. The Carnegie-Mellon University Distributed Speech Recognition System. Speech Technology (1986-04)
- (2) F.Alleva, R.Bisiani, A.Forin, R.Lerner. A Distributed System Architecture for Speech Recognition. ICASSP'86[1986-04]
- (3) Y.L.Chow, R.Schwartz, S.Roucos, O.Kimball, P.Price, F.Kubala, M.O.Dunham, M.Krasner, J.Makhoul. The Role of Word-Dependent Coarticulatory Recognition System. ICASSP'86 (1986-04)
- (4) CMU Speech Group, Summary of Results of the Five-Year Research Effort at Carnegie-Mellon University. Technical Report, Computer Science Department, Carnegie-Mellon University, (1977)
- (5) CMU Speech Group, Raj Reddy. A Continuous Speech Recognition System, Volume I: Technical Proposal, Carnegie-Mellon University (1984-05)
- (6) R.A.Cole, R.M.Stern, M.S.Philips, S.M.Brill, P.Specker, A.P.Pilant. Feature-Based Speaker-Independent Recognition of English Letters. ICASSP'83,(1983-04)
- (7) R.A.Cole. Phonetic Classification in New Generation Speech Recognition Systems, Speech Tech'86(1986-04)
- (8) J.P. Haton. Knowledge-Based and Expert System in Automatic Speech Recognition. New System and Architecture for Automatic Speech Recognition and Synthesis, Springer-Verlag (1984)
- (9) 川端、好田. トップダウン音素照合を用いるワードスポッティング. 音声研究会資料 SP86-24 (1986-07)
- (10) D.H.Klatt. Review of the ARPA Speech Understanding Project. JASA,62,(1977-06)

表5-2 Trigram モデルによる誤り訂正の例

[input sentence]

I will be in touch with you very soon to make an appointment to visit your office period
[recognized candidates without trigram model] 3 errors

[1st] by will be in touch with you very soon to me an appointment to visit york office period

[2nd] I no the it much if new area some two the am important two accept your benefits area

[3rd] my role me am get proof do period seven do male them invite do set or obious very

[recognition results after trigram model] 0 error

I will be in touch with you very soon to make an appointment to visit your office period

[input sentence]

but I find it much more natural and I submit it for your consideration period
[recognized candidates without trigram model] 4 errors

[1st] but I find it much or natural them I say it for york consideration period

[2nd] might by kind that touch more naturally and by submit that far your exclam-point area

[3rd] got my from did not award matter am like site did more work concern very

[recognition results after trigram model] 1 error

but I find it much more natural and I say it for your consideration period

[input sentence]

I am looking forward to seeing you soon period

[recognized candidates without trigram model] 3 errors

[1st] by an living forward to seeing you soon period

[2nd] I am looking far two see in some very

[3rd] my in into for can city do seven area

[recognition results after trigram model] 2 errors

I am living forward to see you soon period

[input sentence]

on the other hand we do not realize all the advantages of living in a city period
[recognized candidates without trigram model] 6 errors

[1st] all the about and me to what realize all the advantages of living in a city period

[2nd] on be upper am we do would arrived role me applicants a looking it the seven area

[3rd] our me other an make two but almost will be positions up within a make submit direct

[recognition results after trigram model] 2 errors

all the other and we do not realize all the advantages of living in a city period

- (11) 好田、中津、鹿野、伊藤. 音声によるオンライン質問回答システム. 音響学会誌、34-3、(1978-03)
- (12) H.C.Leung, V.W.Zue. Visual Characterization of Speech Spectrogram. ICASSP'86 (1986-04)
- (13) B.T.Lowerre. The Harpy Speech Recognition System. PhD thesis, Computer Science Department, Carnegie-Mellon University, (1976)
- (14) 松永、鹿野、Top-Down音韻認識とBottom-Up音韻認識を融合した音声認識. 信学論(D),J68-D,9 (1985-09)
- (15) Textbook of MIT Spectrogram Reading Course. spectrogram reading seminar in ATR (1987-01)
- (16) Sei-ichi Nakagawa. A Machine Understanding System for Spoken Japanese Sentences, 京都大学博士論文(1976-10)
- (17) 中津、好田. 会話音声の機械認識における音響処理、信学論(D),J61-D,4,(1978-04)
- (18) 新美、小林、浅見、三木. 「SPOKEN BASIC」の認識システム、情報処理、5、(1977-05)
- (19) 尾関和彦. 最適文節列を選択するための多段決定アルゴリズム. 音声研究会資料、SP86-32 (1986-07)
- (20) L.R.Rabiner, B.H.Juang. An Introduction to Hidden Markov Model, IEEE ASSP Magazine Vol.3,No,1 (1986-01)

- (21) 関口、重永. 日本語文章の音声認識システム. 音響学会誌、34-3、(1978-03)
- (22) 鹿野、好田. 会話音声の機械認識における言語処理、信学論(D),J61-D,4,(1978-04)
- (23) 鹿野清宏. 会話音声自動認識システムに関する研究. 名古屋大学博士論文(1981)
- (24) Kiyohiro Shikano, K. Lee, Raj Reddy. Speaker Adaptation through Vector Quantization. ICASSP'86 (1986-04),あるいは、Technical Report, Computer Science Department, Carnegie-Mellon University, (1987-02)
- (25) K.Shikano. Improvement of Word Recognition Results by Trigram Model. ICASSP'87 (1987-04)
- (26) V.W. Zue. The Use of Speech Knowledge in Automatic Speech Recognition. Proceedings of the IEEE. Special Issue on Man-Machine Speech Communication. (1985-11)
- (27) V.W.Zue, L.F.Lamel. An Expert Spectrogram Reader: A Knowledge-Based Approach to Speech Recognition. ICASSP'86 (1986-04)
- (28) Alexander Waibel. Prosodics in Speech Recognition, Carnegie-Mellon University PhD thesis (1986-07)
- (29) Brown University, Brown Corpus, Technical Report, Brown University, 1967
- (30) IBM speech group, A Real-Time, Isolated-Word, Speech Recognition System for Dictation Transcription, ICASSP'85, (1985-03)
- (31) F.Jelinek, The Development of an Experimental Discrete Dictation Recognizer, Proceedings of the IEEE, Vol.73, No.11, (1985-11)
- (32) F.Jelinek, Continuous Speech Recognition by Statistical Methods, Proceedings of the IEEE, Vol.64, No.4, (1976-04)
- (33) 鹿野清宏, 入力音声のベクトル量子化による単語音声認識, 音声研究会資料, S82-60, pp.473-480, (1982-12)
- (34) 鹿野、相川, Staggered Array DP マッチング, 音声研究会資料, S82-15, pp.113-120, (1982-06)
- (35) K.Shikano, Text-Independent Speaker Recognition Experiments using Codebooks in Vector Quantization, Tech. Report, CMU-CS-86-101, Carnegie-Mellon University, (1986-02)
- (36) 鹿野、樽松、音声理解研究の動向、音響学会誌4-2巻12号、(1986-12)
- (37) 杉山、鹿野、ピークに重みをおいたLPCスペクトルマッチング尺度、信学論(A),J64-A,5,(1981-05)
- (38) 杉山、鹿野、周波数軸重み付けLPCスペクトルマッチング尺度、信学論(A),J65-A,9,(1982-09)
- (39) S.Furui, Speaker-Independent Isolated Word Recognition Using Dynamic Features of Speech Spectrum, IEEE tr. ASSP, Vol. ASSP-34, No.1,(1986-02)
- (40) K.Shikano, Evaluation of LPC Spectral Matching Measures for Phonetic Unit Recognition, Technical Report, Computer Science Department, Carnegie-Mellon University, CMU-CS-108 (1986-02)