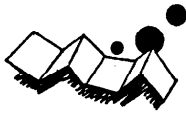


解説



人工知能への期待†

坂井 利之††

1. はしがき

人工知能研究の動向や紹介は今までに幾度となく行われた。ここでは在来の形式や観点にとらわれることなく、人工知能の根底に存在する理念は何であろうか、そしてその手法はどのようなものであるかを考えてみたい。また人工知能研究は、一部理論家の興味本位のものであるとか、あるいは特殊な分野では実験を伴って研究されているけれども、通信工学、情報工学の現実的課題とは遊離したものであって、現段階では直接的に結びつき難いのではないかという懸念も持たれている。

そこで、ここでは、人工知能の考え方や手法を導入すると、通信伝送、情報処理において、どのように質的、量的な変貌をきたす可能性があるかを論じたい。

情報源からのシンボルの出現確率に基礎をおくシャノンの通信理論、あるいは個々のメッセージや情報を画一的に扱う情報処理の立場や手法ではない。情報源の根源的構造に関する先験的知識を利用して、現に通信や処理の対象となっているメッセージに人工知能的手法を展開し、局所的で、かつ限定時間内に情報処理を行う具体例について述べよう。

これらの具体例は、積み木の世界 (block world) といったように、通信工学や情報処理の現実のメイン・テーマと遊離したものではない。たとえば文書パターンのファクシミリ伝送や計算機入力を前提とした文書パターンの構造解析とか、あるいはマイクロホンの設置場所が音声情報源から隔離している場合の SN 比改善などといった通信工学上の基本的考察を必要とする問題である。さらに音声認識や文字読取などの情報処理において、音声とか文字などの先験的な情報源知識を導入することの有効性とか、情報源の構造に適應した処理手順(処理アルゴリズムの優先順位の設定など)

の採用の有効性などについても言及したい。

これらの試みは、人工知能的手法のチャレンジ宣言であるともいえる。すなわち通信工学技術者、情報処理技術者が在来の観点から検討し、真剣に研究に取り組んでいる中心的課題に対して、人工知能的思考を適用するとその結果がかなり良くなるかどうかで、人工知能的手法の真価を問おうとするものである。

情報工学の学問的、技術的存在の証明を、一つは人工知能で明らかにしたいと考えているのである。

情報工学の主柱として期待するものは、単に人工知能 (artificial intelligence) だけでなく、物理的 (ハードウェア) 制限を超越して論理的に扱う virtual の思想と技術、モジュールやブロックの接合に不可欠な interface の定義技術などがあるが、さらにここでもふれるモデルやコミュニケーション思想と手法もきわめて重要であると思っている。

2. 人工知能的手法とは何か

符号表現のシンボル系列、音声波形や音声信号パターン、文字や記号・図・表・写真などの混じった文書パターンなどは、それぞれ情報媒体は異なっているが、人間は情報を伝達あるいは蓄積するためのメッセージとして理解している。ここではメッセージという言葉を広義に上述のような一次元、二次元のアナログまたはデジタルの信号やパターンと考えよう。

人間の行っている情報の授受、コンピュータの扱っている情報処理は、いずれも個々のメッセージごとに効果のあらわれる具体的なコミュニケーションや処理である。多数のメッセージに対して統計的に算出される統計量や総合的な判断ではない。

したがって人工知能の手法を、現実的な通信工学、情報処理工学の課題の上に展開しようとするれば、次のような条件が必要となる。

(1) 個々のメッセージを処理できること

(2) 個々のメッセージに対する評価関数が存在すること (例: 類似度, 手続的優先順位, 情報源の類

† Anticipative Applications of AI Methodology by Toshiyuki SAKAI (Dept. of Information Science, Kyoto University).

†† 京都大学工学部情報工学科

別)

(3) 二者間のコミュニケーションが可能なこと
これらをさらにキーワードに分解し、項目を整理すると次のようになるが、その詳細な内容については末尾の文献を参照したり、以後の説明で逐次明確になるだろう。

(a) 限定時間処理……局所メッセージ処理；探索空間・時間の節約法；処理誤り対策

(イ) 汎用メッセージ処理法……bottom up 手法^{1),2)}で入力メッセージ処理を行い内容理解の階層を上げていく。

(ロ) タスクに依存した処理法……メッセージを生成する情報源に関する拘束条件やメッセージ目的を明確にした、タスクの設定に由来する top down 手法^{1),2)}；効果としては、モデルによる予測が階層の低位レベルに適切に働き情報源の先験的知識の活用で物理量（たとえば SN 比）に関する条件がゆるめられる。

(ハ) ハイブリッド処理……入力メッセージに対する局所メッセージ処理結果の活用 (bottom up 情報の知識化) と top down 情報による選択的、(探索空間節約) 階層的な入力メッセージの処理³⁾による、処理の効率化、高速化、精密化が可能。

(b) 評価関数による処理

(イ) タスクに依存する評価関数……処理対象、処理目的を明確にし、処理装置の知識、能力を加味して決定可能となる。したがってタスクが混在しているときは、まず類別処理が必要となる。

(ロ) 個々のメッセージに対する評価値割付けをメッセージのコンテキストの中で行う。

(ハ) 評価値は広義の二者比較で行う。
低位階層での比較…入力メッセージに対して
高位階層での比較…モデル²⁾、ルール²⁾に対して
ここでは生成法則、推論法則が有効である。

(ニ) 二者比較に先立ってマッチングや写像変換、学習などの処理が必要。

(c) コミュニケーション機能の必要性

(イ) 情報処理を希望する人 (マン) と情報処理をするコンピュータ (マシン) との間で必要となる。使用する場面の例はモデル構築 (変更)^{1),2)}、パラメータ指定²⁾ などである。

(ロ) 階層的に同レベルでの二者の比較、評価値の割当てのための手順の場合とか、階層的にレベル

の異なる二者間でのコミュニケーション (相互干渉, 比較, 変換) 作業のための手順遂行をさせるときに必要となる。

たとえば、類似度検定、優先手順の選択とか top down 手法 (モデル, ルール, 判定結果) からの低位メッセージ予測 (生成), 学習指示⁴⁾ ならびに bottom up 手法 (入力メッセージ) からのシンボル化・ルール化などの場面である。

3. 通信伝送・情報処理を考え直そう

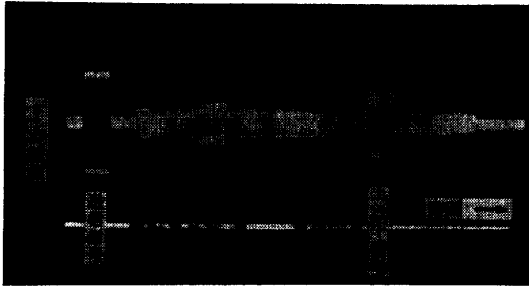
3.1 先験知識と SN 比

シャノンの通信理論では、周知のように先験的知識として情報源シンボル (メッセージ) の出現確率を用い、符号化、通信路容量、SN 比などの諸量の算出とか関係式の導出をしている。

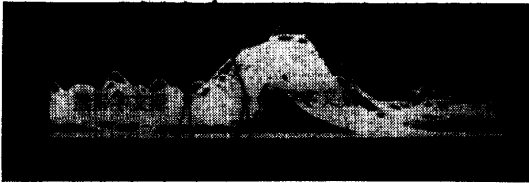
ところが多数のメッセージの統計量として、集約された知識だけでなく、そのメッセージを生成する情報源の構造的知識を知って処理を展開する*と、どう変わるだろうか。まず実例としてノイズに埋もれたレーダ・エコー信号の検出を考えよう。エコー信号の場合には発射送信パルスの時刻を知っている。その時刻から測って、対象物体で反射されたエコー信号が発射地点まで返ってくる時間を計測している。この何回かの繰返しパルス時間 (パルス周期内のエコーメッセージの数カ分に対する限定時間処理で、今の場合には積分時間) 対象物体は動いていないと近似できるという対象世界の知識を利用する。エコー信号は発射送信パルスより一定の時間遅れの位置に定常的に出現するが、ノイズはランダムな位置に生ずる。したがって一定期間積分するとエコー信号のエネルギーは着実に累積されるのに反し、ノイズは累積されない。したがって一定時間の積分により SN 比が改善され、ノイズに埋もれた信号が検出できる。(図-1 参照)

次にスペクトル拡散通信方式の内、周波数ホッピング方式について考える。スペクトル拡散通信方式は近時注目を集めている方式で、共通の媒体空間を利用し、多くのペア間の通信を行うが、混信をなくし、SN 比を改善するのに有望な方式である。その内、周波数ホッピング方式は、信号伝送に必要な周波数帯域より著しく広帯域の周波数を利用し、瞬時信号周波数をその範囲内で符号により跳躍させるものである。知識として、送信搬送波の瞬時周波数の跳躍割当ての符号語ルールを知っている場合 (正しい通信ペア) と、それを知らない場合との SN 比の優劣が問題になる。送

* 人工知能的な解釈を施すことに対応



(a) 送信信号とエコー信号



(b) エコー信号部の拡大

図-1 流星レーダ・エコーの積分

信号符号語（メッセージ）が n 個のパルス信号（ビット信号）の時系列で表現されるとすると、受信側では知識として知っている通り、送信信号の含まれている瞬時跳躍搬送波帯域を、逐次追跡すれば信号の脱落による signal の劣化はない。これに反し、跳躍周波数帯域の知識を持っていない受信者は、 n 回の受信で解読されるメッセージ信号の内、遙かに数少ない受信信号のみからメッセージを推定しなければならない。この場合はメッセージ信号 S の劣化ははなはだしい。スペクトル拡散通信方式の拡散知識を持っている受信者は、有限時間（ n パルス期間）内の符号化処理中に S の劣化は少なく N の相対的減少* と相俟って、 SN 比が改善される。拡散知識を持っていない受信者は、 S の劣化がはなはだしく、 SN 比の著しく悪い受信状態になってしまう。

上例 2 つでは有限一定時間内における情報源の事実または約束としての知識を利用して SN 比改善をはかっている。

これをもう少し一般化して、情報源の空間的構造知識、すなわち情報源の世界の構造に基づく境界条件、信号生成の際の拘束条件を知ることが、入力メッセー

* 非白色雑音の場合、雑音周波数帯域に空間的に合致する累積確率が減少する。持続性雑音の場合、雑音周波数帯域に時間的に合致する累積確率が減少する。

** 送信者のみの条件を考えるのが、“コ”の空間……例：放送、講演。送受信、受信者の二者をシステム要素と考えるのが、“ソ”の空間……例：室内電話系、対談。送信者、受信者、この両者の外に環境、伝送系もアクティブなシステム要素と考えるのが“ア”の空間……例：会議電話、街頭録音、討論会。

ジの処理においていかに貢献するかという、知識と物理量（ SN 比）との交換可能性を調べることは興味あるテーマである。

レーダ・エコーの検出例では、対象物体の移動速度とレーダ機器・対象物体間の距離とで定常性の仮定は変るが、実用的な範囲では、定常性は完璧である。ところが音声のように、共鳴に基づく生成モデルの情報源に対しては、どの程度定常性仮定が成立するだろうか。そして音声信号と非定常またはランダムなノイズとの混在が不可避な空間、換言すれば送信者、受信者、媒体空間が対等に考慮されなければならない“ア”の**空間⁵⁾では、どのように知識と SN 比のトレード・オフが行われるかは面白いテーマである。私が“音声情報フィルタ”と呼ぶものは、音声という情報源と、非音声というノイズ情報源との構造的差異の知識を利用して SN 比改善をはかろうというものである。

3.2 知識化のための人工知能的手法

情報源のもっている構造の信号メッセージ内での表現は、再現性という形で把握できる。信頼できるものは、多くの個々のメッセージに対して共通に出現したり、抽出できる性質 (property) である。property や relation を見出すためには、入力メッセージから、タスクに依存しない汎用的な bottom up の手法³⁾によって、特徴パラメータ、あるいはシンボルとして、構造記述の階層を高位へ上げていく、汎用メッセージ処理法が必要である。

情報源の構造知識は、最も簡単なものから順に書く」と統計的知識、先験的知識、発見的知識となるであろう。

先験的知識というのは情報源の構造モデルを、確率、統計として数値のみで把握するのではなく、より構造記述的に、ある性質 (property) の有無、構造記述パラメータ間の関係 (relation) として先験的に与えられるものである。文献 2) と本稿 4.5 節参照。

発見的というのは、先験的知識を、今現に処理している局所メッセージの bottom up 処理によって、変更可能であると考え、さらには、その構造記述の変更にとどまらず、局所メッセージ処理のアルゴリズム自体の選択、手順の変更や top down 手法にも影響を与えることができるというもので、一般的にはハイブリッド処理といえよう。ハイブリッド処理は、bottom up と top down の両手法がループを形成して³⁾、その停止は評価値によって決定される⁶⁾。

人工知能的手法の中でも、選択あるいは探索ルール

は研究テーマとしてきわめて大きい位置づけをもっている。すなわち組合せ論理や可能性として考えれば天文学的な数字に達するものを全数チェックしている。時間的にも空間的にも到底及ばない。そこで知識や評価値によって範囲を限定させる²⁾探索アルゴリズム、評価値によるマッチング候補の優先順位順序づけやカットオフなどである。

また知識には構造的に記述されるものと、手続的に記述されるものがある。このいずれを選ぶかは、ハードウェアの技術の進歩によって、大きく変貌していくと考えられる。

4. 画像での具体的な例証

4.1 文書パターンの構造把握による領域抽出

文書パターンから、写真領域、文字領域、グラフ領域という構造情報の抽出を行い、それらを簡単な矩形として記述する方法について調べてみた⁷⁾。その目的は、文書パターンの伝送、蓄積、検索における冗長度圧縮、領域の選別による情報の分類、希望しない情報の排除などを可能とする柔軟な検索システムを構成することである。文字領域については、さらに、文字の縦書き、横書き、および、その行を抽出することも行った。

一般に文書画像が与えられたときに、その画像上に一様に情報が分布していることはほとんどない。すなわち、入力画像には空間的に情報が偏在している。この情報の偏在は、文書パターンの構造に起因すると考えられ、その構造は図-2のような tree で表わされる階層性を持っている。我々が採用した構造情報抽出処理の方法と特徴は次のとおりである。

i) 多値入力画像を、ディザ法、および単一閾値によって各々2値化する。前者は、写真領域抽出用の2値画像、後者は、文字領域、グラフ領域など線図形領域抽出用の2値画像を生成する。以後の処理はすべて2値画像上でやっているが、これにより処理の単純化、高速化が図られるからである。

ii) 2値画像を、ランレングス (run-length) 符号化し、文書パターンをランレングスに関するマルコフ情報源とみて、写真領域、グラフ領域、文字領域の判定を行う。この方法を用いることにより、処理の単純化、高速化と共に、従来の伝送・蓄積方法との整合性が図られ、ハードウェア化が容易となる。

iii) 抽出された領域は、ラスタスキャンの方向を1つの辺とする簡単な矩形で記述する。これは、処理の

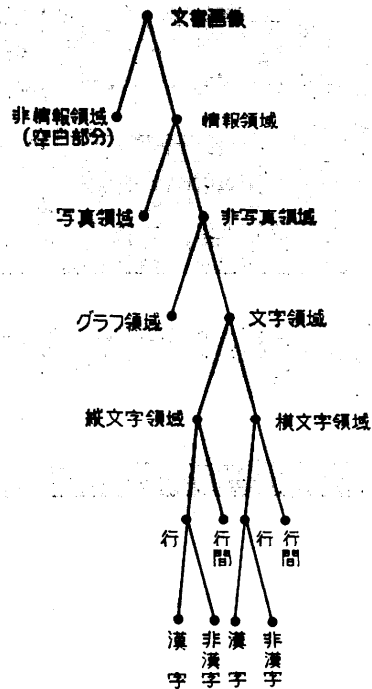


図-2 文書画像の構造情報の階層性

単純化と共に、処理結果を冗長度圧縮に利用する際、構造情報が複雑になると、構造情報自体の情報量が大きくなってしまいうため、できるだけ簡単な記述が要求されるからである。

iv) 領域の抽出結果は矩形で記述されるため、高い精度を必要としない。したがって、処理の中間結果は縮小画像上に記録し、この縮小画像上で矩形の抽出処理が行われる。

ランレングスを利用した領域抽出法で、文字・写真・グラフ領域はそれぞれ次のような特徴になる。すなわち、文字領域における白ランは、比較的短いものが多く、グラフ領域における白ランは、比較的長いものが多い。また、黒ランは、いずれの領域でも、線図形という構造のため短いものが多いと考えられる。したがって、白ラン、黒ラン共に短いランばかり観測されれば、その部分は文字が群をなしていると考え、それを基にして文字領域を抽出することができる。

次に文字領域を例にとって要点のみ説明する。(図-3参照、番号は本文に対応)

① 2値画像をラスタスキャンしてランレングスに直す。

② 白ランレングスを2重マルコフ情報源として観

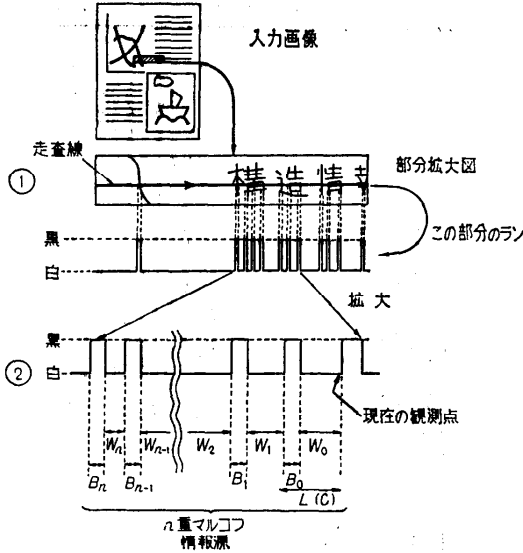


図-3 ランレングスを利用した領域抽出法

測し、現在の観測点までの3個の白ランレングス(W_2, W_1, W_0)がすべて δ (≈ 32) より小さいとき、現在の観測点の直前の黒ラン (B_0)、白ラン (W_0) の組 l を文字領域 $L(c)$ の要素と仮判定する。

③ 雑音処理後、要素をまとまりのよい矩形で囲む。

図-4 に③処理後の結果の一例を示す。

文字領域は、それ自体、文書パターンの構成要素の1つであるが、この領域はさらに、行と行間に分けることができる。この行の抽出を行うことにより次の情報処理としては、文字列の一次元的認識(OCR)のみでよいとか、特徴パラメータによる検索*を可能にする前段処理になっているともいえる。すなわち文書パターンの構造把握は、通信伝送、情報蓄積、その利用という面から考えても必要であり、これを実施するには写真領域をまず探し、次に線図形……という構造情報(図-2)に起因する処理手順の階層性(優先順位)が存在することがわかる。

4.2 線図形の SN 比改善

上述の領域抽出の例でもわかるように濃淡写真と線図形の構造は異なった構造モデルで記述されている。それでは線図形という構造はどのように(先験的)知識として記述できるか。またこの知識を用いると、“音声情報フィルタ”として述べたように“文書のコピーのコピー……”のように、SN 比が悪化している“ア”

* 詳細は、近く発表予定。

の空間で、オリジナルの文書に近い清書が復元可能となるかどうかは興味あるテーマである。

文書パターンのファクシミリ伝送あるいはコピー作成は、オリジナルの文書が送信原稿であり、途中の通信路としては、ファクシミリ伝送中の電気ノイズ、コピー作成中の電気ノイズ、紙の質と光学系、電気系に起因するノイズ源があって、かなり品質の悪い通信路と考えられる。したがって SN 比の改善はコピーの場合には特に重要である。線図形の構造知識を先験的に与えて処理する場合のシミュレーションの結果を図-5 に示す。

図-5(a)はオリジナルで、線図形としての信号とノイズが共存しているものと、その部分拡大図を示した写真、(b)は文字の構造情報を与えて SN 比改善を行った結果である。コンピュータによるシミュレーションは限定時間内に、しかも実際のファクシミリ、あるいはコピー機器に容易に適用可能なよう配慮した方式*で行った。

5. 音声での具体的な例証

5.1 情報源の鼻音化の有無の類別と音声認識

人間の音声は、母音や有声子音のときは肺からの呼気流によって声帯が振動し、その高調波が口腔、鼻腔などの共鳴、反共鳴によって音声波帯域のスペクトル包絡線が決定され、特定の音韻の音声が発生される。その際同じ母音の場合でも、口腔以外に鼻腔が結合されると鼻音化母音となる。たとえば日本語の /ichi/ の最初の母音 [i] と鼻音化される /ni/ の [i] との差異は大きい。このように同じ音韻でも単語とか文章の中に出現するときは、前後の音韻の影響(コンテキスト)を受ける調音結合の現象が不可避となる。

そこで音韻を認識する場合機械の中に蓄える標準スペクトルパターン(今の例では簡単に母音として話をする)と、発声された入力音声のスペクトルパターンとの類似性のマッチング度を調べることになるが、鼻音化の有無にかかわらず、ただ一つの標準パターンを設けるだけですませるのは、果して妥当かどうかの疑問が生ずる。

そこで日本語 5 母音のそれぞれに、鼻音化されたものとされないものの 2 つのクラス(合計 10 の標準パターン)を用意できるようにした。そしてそれらをより特定個人の標準パターンにふさわしいように、教師なしの学習を行い、情報源の構造により忠実に適応的にマッチするようにすると、認識結果が向上することを

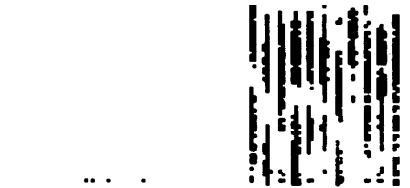


図4 (a) オリジナル文書画像 (多値画像)

この図は、縦文字領域抽出処理の一例を示している。元の文書画像から、縦文字領域を抽出し、縮小表示している。抽出された領域は、縦方向に長い黒い帯状の領域として表示されている。



(a) オリジナル文書画像 (多値画像)



(c) 文字領域の縮小画像



(d) 文字領域の抽出 (矩形表示)

図4 縦文字領域抽出処理例

確めた⁴⁾。

その処理の概要と認識結果について簡単に述べる。単語音声認識の流れ図は図-6に示してあるが、本実験

では、母音の標準スペクトルパターンは識別を開始する以前から2つのクラスに分類しているのではなく、教師なしによる学習を行っていく段階において自動的

化を多く含む領域は、このようにして抽出される。抽出された領域は、縦方向に長い黒い帯状の領域として表示されている。この領域は、縦文字領域抽出処理の結果として抽出された領域である。

図4 (a) オリジナル文書画像 (多値画像)

この図は、縦文字領域抽出処理の一例を示している。元の文書画像から、縦文字領域を抽出し、縮小表示している。抽出された領域は、縦方向に長い黒い帯状の領域として表示されている。



(a) オリジナル文書画像 (多値画像)



(c) 文字領域の縮小画像



(d) 文字領域の抽出 (矩形表示)

図4 縦文字領域抽出処理例

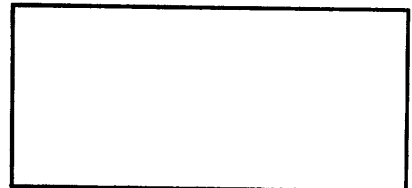
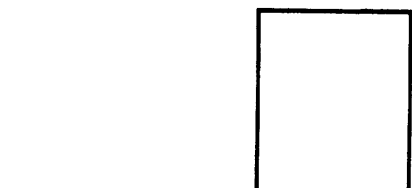
この図は、縦文字領域抽出処理の一例を示している。元の文書画像から、縦文字領域を抽出し、縮小表示している。抽出された領域は、縦方向に長い黒い帯状の領域として表示されている。

図4 (a) オリジナル文書画像 (多値画像)

この図は、縦文字領域抽出処理の一例を示している。元の文書画像から、縦文字領域を抽出し、縮小表示している。抽出された領域は、縦方向に長い黒い帯状の領域として表示されている。



(b) 写真領域削除後の結果 (2値画像)

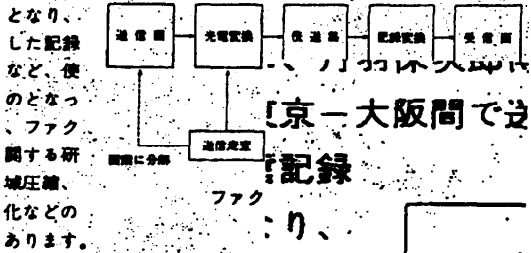


(d) 文字領域の抽出 (矩形表示)

図4 縦文字領域抽出処理例

はこれを電気的手段によって実現し、その基本構成は図のよう
 。送信側は走査によって画素に分解され、電気信号に変換され
 ます。受信側では、送られてきた電気信号を記録に通した信号
 ながら画素を組み立て、受信画を得ます。

年に、丹羽保次郎博士らによって写真電送機が初めて完成され、
 が東京一大阪間で送られました。その後、電子技術の発達に
 受信記録



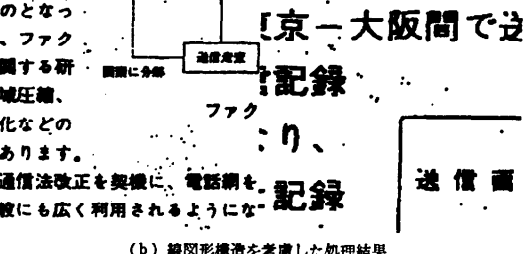
となり、した記録など、使のどなっ、ファク
 に関する研
 城圧縮、
 化などの
 あります。

通信法改正を契機に、電話網を、
 般にも広く利用されるようにな

(a) ファクシミリテストチャート No. 4 に静電複写段階でノイズを
 加えたもの

はこれを電気的手段によって実現し、その基本構成は図のよう
 。送信側は走査によって画素に分解され、電気信号に変換され
 ます。受信側では、送られてきた電気信号を記録に通した信号
 ながら画素を組み立て、受信画を得ます。

年に、丹羽保次郎博士らによって写真電送機が初めて完成され、
 が東京一大阪間で送られました。その後、電子技術の発達に
 受信記録



となり、した記録など、使のどなっ、ファク
 に関する研
 城圧縮、
 化などの
 あります。

通信法改正を契機に、電話網を、
 般にも広く利用されるようにな

(b) 線図形構造を考慮した処理結果

図-5 線図形の SN 比改善例

に分類する方法を用いている。母音スペクトルの2ク
 ラス分類を単に音韻識別の低レベルのみにとどめ、よ
 り高いレベルの標準スペクトルおよび単語辞書として
 はあらかじめ細分化していないのは次の理由による。
 すなわち、どの話者においても、各母音を必ず2つの
 クラスに分類する必要があるかという疑問がある。ある
 コンテキスト中の母音の2つのクラスへの分かれ方は、
 どの話者についても同じであるかという疑問がある。
 標準スペクトルパターンおよび単語辞書を作成する
 ときに、各母音について2つのクラスに分類するた

めには組織だった方法（ルール）が必要となるが、こ
 のためには、あらゆるコンテキストについて検討する
 必要があり困難である。

また教師なしで学習を行っていく上では、誤って学
 習することを極力避けなければならないので、本実験
 では次の3つの方法を用いる。

- ① 音韻識別段階で、母音スペクトルパターンを各
 母音について2クラスに分類する。
- ② 音韻識別結果の信頼度の高いものだけを学習対
 象とする。
- ③ 学習するかどうかの決定に、入力単語音声の識
 別結果を用いる。

教師なしによる学習は、図-6 に示したように次の3
 ステップによって行われる。

ステップ 1: ①入力単語音声中、母音として識別さ
 れたフレーム（短時間たとえば 10 ms のスペクトル・
 パターン）に対し、第1候補母音の信頼度がしきい
 値 (α) 以上であること。

②同一母音として識別されたフレームがしきい値
 (β) 以上連続すること。
 の2つの条件を満足するものを学習候補データとす
 る。

ステップ 2: 入力単語音声がある単語として識別さ
 れたときのマッチング・スコアが、しきい値 (γ) 以上
 であることが以後の学習に採用される試料としての必
 要条件である。ここで第1候補（識別結果に相当）と
 第2候補のマッチング・スコアの差を Δ とすると、

- ① $\Delta \geq$ しきい値 (δ) のときは、第1候補の学習
 用単語辞書に含まれる母音を、
- ② $\Delta <$ しきい値 (δ) のときは、第1候補と第2
 候補の学習用単語辞書に共通に含まれる母音を、学習
 に用いられる母音の種類とする。

ステップ 3: ステップ 1 で得られた学習用データの
 うち、ステップ 2 と矛盾しないものが学習に使われ
 る。このデータは母音 10 (各母音 2 クラス) のスペク
 トルパターンと比較され最も類似しているクラスの学
 習に用いられる。

40 名の男性話者が 10 数字を、各 5 回発声した試料
 について、学習を行わないで識別実験を行った場合に
 識別が困難であった 5 名について次の 3 つの実験を行
 った。

- ① 各母音スペクトルパターンを分類しないで学習
 を行う。
- ② 各母音スペクトルパターンを 2 クラスに分類

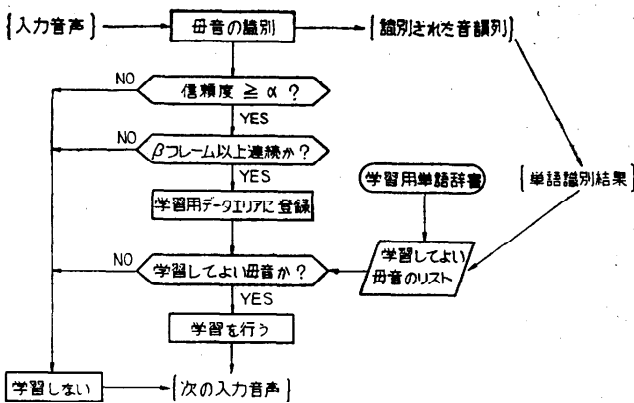


図-6 単語音声認識のシステムと学習用データ決定手続き

表-1 学習による標準スペクトルの2クラス類別と数字音声の識別率との関係

方法	学習なし	①	②	③
話者				
NK	80%	82%	84%	92%
FJ	84	78	92	96
NG	74	82	92	92
TM	90	72	84	86
MT	88	88	92	98
平均	83.2	80.4	88.8	92.8

男性 40 名のうち識別率が悪い 5 名による実験

①日本語母音 5 を学習 ②日本語母音 10 のうち 5 のみ学習

③日本語母音 10 を学習

し、安全のために一方のクラスは全話者共通のスペクトルパターンを残し、他方のクラスのみ学習を行う。

③ 各母音スペクトルパターンを2クラスに分類しながら学習を行う。

参考のために、学習しないで全話者共通のパターンを用いて識別を行った場合もあわせてその実験結果を表-1 に示す。方法①は、学習(提示)順序に一番大きな影響を受け、学習しない場合よりも悪くなることもある。各母音について標準スペクトルパターンのクラスを2種類もつように学習してゆくことが有効であるといえる。

5.2 情報源の構造差の類別と音声認識

ここでは、音声理解システムを考えると、重要となる項目の一つとして、“情報源に対する知識の獲得—情報フィルタの効用”のもつ意味を考えてみよう。

我々人間は、声を聞いたとき、男の声、女の声、子供の声ということがわかる。ところが、音声理解システム^{2),5)}は音声波形を発声過程の逆向きにさかのぼっ

て、音声メッセージを理解することであるが、通常音声理解システムでは、人間の発声器官の大きさを、個人差のハードウェアの変動要因として扱っている。しかし、この音声波の情報源の大きさ、構造を、情報発生源の世界(人工知能的)構造として把握し、その知識をつかんだ上で、次に音声認識の処理をするのと、通常の個人差要因の扱いとは、どう異なってくるかは興味がある。

もし SN のきわめて悪いところ、すなわち音響ノイズが大きく遠くから音声波をマイクで収録せざるを得ないようなときには、特に人間の声らしいものと、

人間の声らしくないものとを分離できたら便利であるとは誰も考える。つまり、情報の内容によって、情報をフィルタできることが、以後の処理に先立って可能ならば、きわめて望ましいことであることがわかる。このように情報の内容、人間の声らしいこと、男の声らしいこと、子供の声らしいことを本格的処理に先立って、ふるいわけすると、音声の認識処理にも好結果をもたらしそうであることを述べよう。

実験は、日本語母音の認識について行った⁶⁾。小学5年生の男・女、20才前後の男・女、40才以上の男・女、それぞれ20人、計20人×6=120人が5つの日本語母音を3回発声したものを機械による認識対象とした。

人間の声を特徴づけるものとして、声帯の振動数であるピッチ周波数、ならびにホルマント周波数、偏自己相関係数、声道断面積、声道長、滑らかなスペクトルなど、ほとんどすべての特徴パラメータの抽出と利用を試みた。このうち、人間の発声器官の構造と関係が深く、話者の類別(上記6グループ)に役立つものとしてピッチ周波数、スペクトルの傾き、擬似声道断面積比を採用し、さらにこの類別をした上で、母音の認識に用いたパラメータは、擬似声道断面積比(セクション数10)である。

表-2 に実験結果を示す。表-2(a)は、声帯振動数で、子供(男・女)は260 Hz以上、成人女子200 Hz以上、成人男子150 Hz以下であることがわかる。したがって、これら6グループをどのように類別するのが、発声器官の構造と合致するかをいろいろと分類し、またこれらを機械により、ピッチ周波数などでクラスわけを行わせると、スコアは表-2(b)のようにな

表-2 発声器官構造の類別と日本語母音認識率

年齢	11才	~20才	>40才
性別			
男	269	146	139
女	268	246	205

(a) 音声ピッチ周波数 (声帯振動数)
各グループ20人 合計120人
日本語母音5×3回発声

クラスわけ	識別率	クラス	母音	備考
子供,男(成人),女(成人) 3クラス	80.7 79.3 [100=教示]	94.4 91.1 96.4 92.2		◎ 教示効果: 大
男(11),男(20),男(40) 女(11),女(20),女(40) 6クラス	63.8 56.4 最低スコア [100=教示]	94.1 88.0 →スコア急減 98.1 89.6		× クラスわけ無意味?
男 女 2クラス	72.4 70.7 [100=教示]	92.5 90.0 94.7 89.8		△ 教示効果: 無意味 女性スコア寄与
区別せず 1クラス	—	92.6 88.9		
男(成人),女・子供 2クラス	94.6 93.8 最高スコア [100=教示]	94.6 90.7 95.7 90.4		◎ クラスわけしたがって 教示効果: 小

(b) クラスわけと母音識別率 (%)
左の列: 120人話者 [1800発話] (事前分析ならびに本番認識)
右の列: 30人話者 [450発話] (事前分析 90人: 本番認識 30人)
3:1

る。男ということだけで、子供のように 250 Hz 以上もあるものと、150 Hz 以下の成人男性のを一緒にすることの不自然さが、男女2クラスというクラスわけのスコアの72.4%という悪いスコアによく出ている。6クラスというのは、全く情報源の構造内容とは関係のない年齢、性別で便宜的に分類したもので、機械によるクラスわけは、63.8%とクラスわけの5種類の中で最低のスコアとなっている。

各列の中で左側の数字は、発話を標準モデル作成のため事前に全部利用し、同じサンプルを次の認識にも用いたものである。同じ欄で右側の数字は、分析に用いた音声サンプルは各クラス20人×3/4=15人で、分析に用いなかった残りの5人分のサンプルを認識実験に利用した場合のスコアでそれぞれ左側より低くなっている。

さらに各行の欄で、下の方にあるのは、前処理の話

- * クラスわけのスコアは、教示が完全であるから100%であって次段処理の母音識別率のみ、有意な数値となり、教示すると識別率は最高スコアとなる。
- ** 母音識別率で上段と下段とを比較すると、クラスわけを教示した方がよいスコアとなっていて教示効果大きい。逆にいうとクラスわけが有意ともいえる。男女2クラスの場合は表-2の備考にあるようにクラスわけが無意味ともいえる。男(成人),女・子供2クラスの場合は表-2(a),表-2(b)で分かるようにクラスわけのスコアが最高によいので、教示しても母音の識別率は3クラスの場合ほど、向上しない(教示効果;小)。

者クラスわけを機械に教示したものである*。これらから話者を3クラスへ類別すること、すなわち子供(男・女)、成人男性、成人女性の類別が母音の識別率(左側,右側の条件の両方,特に右側)で、最もよいスコア91.1%を示している**。ただこの実験は、まだ人数も不足であり、5母音だけということ、断定的な発言には資料不足であるが、情報源構造の把握と、それに基づく前処理としての情報フィルタリングの後での認識処理が好結果につながるという仮説をサポートしているようである。

6. 論議と考察

5章までの草稿を書きあげて、コピーを配布し、研究室の人々と数時間論議をした。そこで提出された疑問点や問題点を参考のため掲げてみよう。

“人工知能とは”というのに対して専門家、非専門家とりまぜて次のような意見があった。

① 人工知能とは global なものである。与えられたデータ、与えられたセンサや装置でデータ処理するのはその一部分であって、何をデータとすべきか、どんな入力装置で何をとりべきかまで考えるものである。

さらには入力データの質についても、たとえばステレオマイクでとり両耳できくべきだ、などといった要求まで出すのが人工知能である。

② 人工知能とは理論なのか、それとも手法なのか。種々の既存の対象(分野)に人工知能的手法を適用したといっても、当該分野ではその分野固有の所で、少し観点を変えた位に思わないか。そうすると人工知能という総合領域、境界領域は具体化すると狭くなっていく。固有の対象分野がなければ残るのは、理論か手法のみである。

③ アルゴリズムが明らかに書き下せるものまで人工知能と呼ぶのか、やり方が判然としないものを研究対象とするのが人工知能ではないのか。

④ 人工知能的手法に必要な条件は記述されているが、人工知能の研究分野で現在成果を上げている Expert System とか探索法などは、具体的に示されていない。判っている人工知能的手法を利用すれば、何が出来るかを示す必要はないか。

⑤ 人工知能とは一体理論なのか思想なのか。理論はものの考え方の整理、処理の能率化につながり、見通しをよくする。思想というには少し足りないすれば志向ということで、そのような考え方、その方向で

の考察をすることをそれぞれの領域でやってみたらという提言が目的であるのか。“情報マインド”と言う言葉と同様に「“AI マインド”で考えよ」ということなのか。

⑥ 書かれている例証では、すべて人が知識とかアルゴリズムを与えている。全く未知のものに対して、機械が処理できるようにしていくのが本来の人工知能であると思っているが、それに対してはどうか。例証でもやはり人工知能と言えるのか。

これらの問題提起や疑問に対してただちに答えられるものもある。このように考えているという答え方になるものもある。現在鋭意研究中で、いずれははっきりしてくるというものもある。

上述の質問に対してここでは返答を示さない。ただ1つコメントするとすれば、人工知能的手法と的をつけたのは、その手法の重要なもの、あるいは考え方の1つまたは幾つかを利用するという意味であって、全部が全部含まれているという意味ではもちろんない。

具体例としてあげた例題の中で、どこが一体人工知能的で、どこが在来の通信工学、情報処理と異なるのかの明確な説明を書かなければならないと思いながら結局入らなくなった。これを明らかにしないことには、技術的、学問的説得力に欠けると言われても仕方がない。

いずれにしても、この解説(?)文章では不十分である。しかし抽象的な言葉だけでなく幾つかの例題をもって、方向と考え方を示したので、賛成派であって

も反対派でも、あるいは中道派の人に対しても動機づけにはなると思っている。詳細については研究論文として逐次発表されていくので、もう少し動勢が明確になってから(それでは遅すぎる?!)研究や開発の方針・態度を決められるのも一案かとも思われる。

学会誌で、しかも異例の書き方と内容のものになったが、ご寛容を願うものである。

参考文献

- 1) 坂井: 1980年代の画像処理, 情報処理 Vol. 21, No. 6, pp. 639-644 (昭55.6).
- 2) 坂井: 人工知能一特に音声, 画像理解を中心として— bit 11, 6, pp. 597-604 (昭54.6).
- 3) 大田友一, 金出武雄, 坂井: ボトムアップ制御とトップダウン制御の組合せによる領域解析, 情報処理学会論文誌 Vol. 21, No. 2, pp. 116-123 (昭55.3).
- 4) 中川聖一, 坂井: 個人差の種々の学習機能をもつ実時間単語音声識別システム, 電子通信学会論文誌, Vol. J61-D No. 6, pp. 395-402 (昭53.6).
- 5) 坂井: 情報システムにおける音声の認識と合成, 情報処理 Vol. 21, No. 8, pp. 820-828 (昭55.8).
- 6) 坂井: 画像処理技術と社会, 電子通信学会誌, Vol. 60, No. 8, pp. 885-893 (昭52.8).
- 7) 村尾充洋, 坂井: 文書画像における構造情報の抽出, 情報処理学会第21回全国大会, 2H-10 (昭55.5).
- 8) 坂井: 人工知能の最近の進歩, 情報処理学会関西支部“最近のデータベースと人工知能の動向—1980年を迎えて—”, pp. 1-11.

(昭和55年8月12日受付)