

特集 「考証：2001年宇宙の旅」

HAL's Legacy にみる人工知能の 現状と将来

The Present and Future of AI
— Views from HAL's Legacy —

大沢 英一
Ei-Ichi Osawa

公立はこだて未来大学
Future University-Hakodate.
osawa@fun.ac.jp, <http://www.fun.ac.jp/>

Keywords: language understanding, emotion, common sense, mind, collaboration.

1. はじめに

1997年に“HAL's Legacy” [Stork 97] (以下、この本をその邦訳タイトルに従って「HAL伝説」と呼ぶ)と題する本がMIT Pressより出版された。これは、1968年にStanley Kubrickにより映画化されたArthur C. ClarkeのSF小説「2001年宇宙の旅」(以下、「2001年」と略す)に登場する人工知能システムHAL 9000(以下、HALと略す)を現時点の人工知能技術に照らし合わせて科学的・工学的に考証したものである。なお、小説のなかでHALは1997年に製作されたという設定になっており、「HAL伝説」はその誕生年にちなんで出版された。

さて、1968年というとまだ人工知能研究の萌芽期である(現在でもまだ萌芽期かもしれないが)。「2001年」はフィクションであるから実現可能かどうかは別の問題として、HALはさまざまな点において、当時の(そして今日の)人工知能をはるかに超えた能力を有しているように見える。ただしHALは、他の多くのSF映画に見られるような全く荒唐無稽な想像による機械というふうでもなく、「HAL伝説」の編集者であるDavid G. Storkに言わせれば、これは“とてもよく考えぬかれた夢”なのである。実際、映画の製作にあたってKubrickは当時の計算機科学や人工知能を取材しているようだ(例えば、MITにMarvin Minskyを訪ねている)。また、作品発表後、HALは人工知能研究の発展過程においてさまざまな局面でたびたび引用されてきた。原作者であるClarkeは「2001年」は未来予測的な位置づけのものではないと主張しているが、人工知能研究のコミュニティの中には、HALを未来の人工知能像、もしくはそれに近いイメージとしてとらえていた研究者が少なからずいたのではないだろうか？

「HAL伝説」では、延べ16人の著名な人工知能研究者がさまざまな視点からHALに関して考察している。

その内容は、小説や映画に描かれたHALが技術的に可能かどうか、それは人工知能があるべき姿として描かれているかどうか、現時点の人工知能技術はHALと比較した場合にどこまで達成されているのか、また、HALのような高度な人工知能システムは人間社会においてどのような社会的位置づけになるのかなど、多岐にわたっている。

本稿では、まず「2001年」のあらすじを紹介し、続いて「HAL伝説」に収録されている論文のうち、何人かの研究者の考え方を、アーキテクチャ、高次知能(常識、心、言語能力、人間との共同作業)、そして感情などに絞って紹介する。

2. 2001年宇宙の旅のあらすじ

「2001年」は人工知能を扱った代表的なSF作品であるから、人工知能に興味を持たれている大半の方々は既にこの物語の内容をご存じであろう。しかしながら、本稿に続く他の論文の都合もあるので、この物語のうち、HALが登場する場面のあらすじを紹介しておくことにする。

『時は2001年、米国の調査チームが月面に異常磁場を発見する。その原因を追求するための極秘プロジェクトが生まれ、異常磁場の中心を掘り起こしたところ、モノリス(直立石)が発見される。このモノリスは精巧に加工された謎の物体で、それが強い磁力を発生していた。さらに驚くことに、モノリス上に堆積した地層から推定したところ、それは400万年前に埋められたものであることがわかった。

あるとき、太陽光を受けたモノリスは木星に向けて強い電波バーストを発信した。科学者たちは、このモノリスは、それが掘り起こされたときにある条件でスイッチが入り電波を発するように仕込まれていたのだらうと推察した。それはつまり、400万年前に高度な科学技術をもった知的な存在が月を訪れていたことを

意味することになる。以上の発見事実や推測は、米国の国家機関と一部の科学者たち以外には極秘とされたが、その電波バーストが向けられた木星に何があるのかを探るために、木星探査プロジェクトが開始された。

モノリスの電波発信より18か月後、宇宙船ディスカバリー1号は木星への航路を進行中であった。このディスカバリー1号こそが、先の木星探査のミッションを担った宇宙船である。ディスカバリー1号の乗組員は5名の人間（うち3名は木星での作業までは人工冬眠状態にあり、残りの2名、ポーマン船長とプール副船長だけが航行任務についている）とHAL 9000という最新式の人工知能システムから構成されていた。HALは画像理解、音声対話、高度な推論といった高次知能をもつ完全無欠なコンピュータであり、それには、宇宙船の航行管理、生命維持装置の管理、人工冬眠中の乗組員の管理、そして乗組員のリクリエーションの相手（HALはチェスの名手である）などの作業が任されていた。

HALは、与えられたミッションに強い情熱を持って取り組み、また自己の能力に自信を持っていた。さらに、乗組員はHALの能力に高い信頼をおいているように見えた。ところが、あるときから乗組員とHALとの関係に徐々に問題が生じ始める。実は、5名の乗組員は自分達に与えられたミッションの本当の目的は知らされておらず（モノリスのことも知らされていない）、それを知っているのはHALだけであった。このことが原因で、HALの心理にある種のジレンマが生じ、HALと乗組員の関係が円滑に進まなくなる。

乗組員の誰も気づかぬうちに生じたHALの複雑な心理状況は、小さな事件を引き起こす。HALはアンテナユニットの異常を感知し、それが72時間以内に故障すると予測した。乗組員は船外活動によりそのユニットを取り外し検査するが異常は発見されない。また、地上にあるHALと同型の他のシステムもユニットは正常であると診断する。このことを契機に、乗組員たちはHALの能力を疑いはじめ、HALの判断ミスによる航行中の事故を予防するために、HALの高次知能を停止しようと企てる。事前にその謀議を察知したHALは、プール副船長を宇宙空間に葬り去り、冬眠中の3名の乗組員を殺害し、ポーマン船長をも殺害しようと試みる。

HALとの格闘のすえ、九死に一生を得たポーマン船長は、HALの高次知能の停止を実行する。この停止作業の過程において、HALは命乞いをし、徐々に薄れる意識のもとで哀れな姿をさらけ出す。

以上が、「2001年」の部分的なあらすじである。実は、上で紹介した部分の前、および後ろに、物語の意図に深く関係する場面があるのだが、その部分にふれるとHALへの焦点がボケるように思われるので、あえて割愛した。

3. 製造可能性

David J. KuckはHALのような計算機を構築できるかどうかに関して考察している。小説版「2001年」では、3段階の飛躍的進歩により、1980年代に入って人工ニューラルネットの自動生成の手法、つまり、人工知能を育てるのに人間の脳の発達と酷似したプロセスが可能になったとしている。確かに1980年代に入ってニューラルネットに対する関心は高まり、比較的単純な要素をネットワークにより多数つなぎ合わせ、トレーニングシーケンスを多数与えることで、特定の機能を学習できるようになった。それにより実用的な成果がいくつか得られたが、飛躍的進歩をもたらしたとは言えないとKuckは述べている。

計算機の性能は、ハードウェアとソフトウェアによって決まる。ハードウェアの性能は、主に集積度とクロック速度により決まる。まず集積度に関して言うと、よく知られたムーアの法則がある。これは大雑把に言うと集積度は毎年2倍に増えるという法則である。実際に、集積回路が作られてから現在まで、集積度はほぼこのくらいのオーダーで上がってきている。クロック速度は、ハードウェアによる遅延を考慮して決定されるが、過去50年を平均すると7年ごとに約10倍になっている。しかし近年では物理的限界に突き当たりつつあり、伸びが鈍化し、最近7年間では2倍にも達していない。ハードウェアの性能向上の概算は以上のとおりであるが、2001年には0.1マイクロルールレベルの集積回路と2nsのクロック速度が可能である。しかしながら、このようなハードウェア技術をもってしてもHALのような計算機を実現するには十分とは言えないとKuckは述べている。

計算機のシステム性能を評価する場合には、ハードウェア技術以外にアーキテクチャとシステムソフトウェアの性能が見逃せない。アーキテクチャに関しては、スケーラブルな並列化がその鍵となってきた。「2001年」ではHALのアーキテクチャについての記述はないが、それについて推察することは可能である。まず、その高性能さ、高機能さから見て並列処理システムであることは間違いない。人間の脳も同時にさまざまな機能を果たすために並列処理を行っていることは明らかである。

未来の計算機は、その物理的な容量に関しては、人間の脳に匹敵することになるかもしれない。しかし、そのような計算機がどのような機能を持つかについては何も暗示しない。重要なのは、脳と相似的なアーキテクチャを人間が与えなくても、AIは実現するかもしれないという可能性を認識しておくことだ。AIと計算機アーキテクチャにおける、重要な飛躍的進歩が何度も起きたとき、今まで以上に強力な推論システムの構

築が可能となるかもしれない。

HALは、さまざまな認識機能や制御機能を実現するために、強力な並列システムを搭載した分散システムであったらう。実際、人間の脳では、多数の類似した独立部分が制御や視覚、聴覚、発声などのさまざまな認識処理を司っており、それぞれの部分は高度に並列化したニューロンの集合体である。したがって、過去30年における計算機科学の発展及び歴史と、今後HALを建造するのに必要な事柄の間には大きな相似点があったのだとKuckは述べている。

4. 常識と心

HALによると、HAL 9000シリーズ・コンピュータは過去に過ちを犯したことがなく、完全無欠でエラーがないことになっている。またアンテナユニットの故障予測の場面においても、過ちを犯すのは人間の常であり、9000シリーズには過去に過ちの記録はないとHALは主張している。常識を備えたプログラムCYCシステムの開発を指揮するDouglas B. Lenatは、HALがどれだけ賢かったのか、また、HALの知性は現在我々が作り出せる知的プログラムの実像と比べてどのようなものであるかを論じている。

Lenatはまず、脳を知識のポンプのようなものだと考える。知識は、さまざまな訓練や経験によって蓄えられる。そのような知識なしには、知的な活動をなし得ない。例えば、話すことを学び、日常世界にひとりで生きていけるようになるためには、たくさんの常識が必要である。常識は共通認識を生み、それによりコミュニケーションの成立を助ける。より効率のよいコミュニケーションをとるためには、共有経験、習慣、共通の専門知識などが頼りになるだろう。現在のコンピュータは常識的知識すら十分にもっておらず、よって人間の発話の大半を理解できないのだとする。

では、HALのようなコンピュータはどうすれば作り出せるのか？ Lenatは次の三つの段階を提案している。

- 何百万という日常の言葉、概念、経験則により常識を形成する。
- 常識の上に自然言語で対話する能力を構築する。それによりさらに知識の基礎を拡大する。
- ある分野で人間の知識の最先端に達したら、実践によりさらにその分野で前進する。

Lenatは自己の推進するCYCプロジェクトにおいて上記の3項目を達成しようと試みている。CYCプロジェクトは1984年にスタートした。まず、CYCに数百万という重要な事実や経験則を与え、知識のポンプに呼び水をさそうと考えた。そして1990年代末までには、それらの豊富な知識をもとに、自然言語での会話や読解によりさらに学習を可能とすることが目標であった。そしてその後は、自動発見の手法でCYCに自己学習さ

せる計画を立てている。

現時点でCYCプロジェクトがどの程度の成功を納めているかに関して、正確な情報は把握していない。後の章で紹介するSchankの考え方と比較すると、CYCの開発過程においては、計算機に経験させるということはどう考えているのか明らかではないが、現在でも常識的知識を与えることを地道に続けているのであろう。

さて、Lenatは、HALは自己の能力を過信してうぬぼれており、愚かだとしている。さらに自己のもったジレンマを解消するために、乗組員の殺害を実行する点などにおいて明らかに常識が欠けているとしている。常識がなければ有意義な形で問題を解決することはできない。HALがもっていたのは付け焼き刃の知性であり、価値観も常識も欠けたために乗組員のほとんど全員が無駄死にさせられる結果につながった。人工知能の開発において、この点は十分に考慮されなければならないとしている。

5. 言語能力

HALは、ほぼ完璧と思われるレベルで言語を使いこなす。HALと同程度に言語を使う計算機は実現可能かどうかについてRoger C. Schankが論じている。

まず、問題解決が知能の本質であるかどうかについて考えている。例えば知能テストは問題解決能力に主眼をおいている。初期の人工知能研究者も同じように考え、チェスを含むあらゆる種類の問題を解くプログラムを構築しようとした。しかし、一定数のルールを憶えることと、状況に応じてそのルールをあてはめられるよう学習すること、さらにこれらのルールをつくり出すことには大きな差がある。学習性が知能の重要な部分なのである。

HALは、物語のさまざまな場面において巧みに言葉を使いこなす。HALは、事前にすべての単語の意味を入力されていたから話せるのか？ そうだとしたら、最初に入力されていた以外のことは理解できない可能性が大きい。言語を使ったさまざまな体験からは何も学習しないのか？

言語と行動を理解するためには、自己の目標、周囲の人々の目標、そしてそれらの優先順位を含めた完結した世界観が必要だ。さらにそれらの目標に関連する方法も、それらを実行するために障害となり得る潜在的問題も、すべて理解しなければならない。HALが製作時点において、ありとあらゆる知識と言語に関する知識をもっていたとすれば、このようなことは理論的には可能だ。しかし、それは問題解決の第一歩にしかならない。なぜなら、知識は定常的ではなく、世界観は新しい経験によりつねに変わっていくからである。HALは、このように自己の経験を通して知識を獲得していくことが可能だったのか？

ここで HAL が人間の感情について語る部分に着目する。物語の中で、HAL は自己の能力に自信をなくしたことはないかと問われる場面がある。それに対して HAL は、自分は過去に間違いを犯したことはなく完全無欠であると答える。さらに HAL は「いらだち」といった感情についても言及するが、HAL が人間の感情について話をするからには、まずその感情を経験している必要がある。HAL は、どのようにしてそれを体験したのか？

HAL は過去に過ちを犯したことがない。つまり自信の欠如した状態を自分では体験したことがないのに、なぜその説明を理解できるのか？

もしかすると自信を欠如した人間に関する話をいくつか聞いて、その状況を分析することで問題点を分析したのかもしれない。しかし、現在の計算機はどうていそんなことはできないし、「2001年」の中にもどのようにしてそれが可能になったのかに関する描写が全くといってない。HAL は 1997 年に製造された時点からある指導者について学習しながら知識を獲得してきたと思わせる場面がある。ある種の知識は教えられることで獲得できるだろうが、感情に関する理解は直接的にせよ間接的にせよ経験して学ぶしかない。このような学習はさまざまな状況における試行錯誤の上で成り立つものだ。

Schank は、自然言語の問題の核心は言語にはないと主張している。研究者が膨大な単語の意味を計算機に与えても、計算機は人間の経験を理解するようにはならない。問題は言語ではなく、知識とその獲得方法にある。つまりいくら高度な計算機があったとしても、人間の子供の場合と同様に、流し込みによる学習はうまくいかないと述べている。

こうなると、HAL のような高度で多様な知識を本当に獲得するためには、あらゆる、つまり膨大な経験を必要とするだろう。そしてそれらの経験の中から学ぶように仕向ける必要がある。しかし、そんなことは現実には不可能だ。映画の中なら効果的であったとしても、現実問題として、計算機に、あらゆる問題に答えられる能力など期待すべきではないし、そんなことは不可能なのだとしている。

「2001年」の HAL に仮定されていた知能のモデルは、人工知能のモデルとして間違っているばかりか、人間知能のモデルとしても間違っていると Schank は主張する。すべての話題を知的にこなせる人間などいるだろうか？ 大半の人間の日常的な非オリジナルな会話というのは、ある種の繰り返しの連続である場合が多い。人間はいつも聞かせたいと思っている話の貯蔵庫であり、他人からの入力に主にそれらの話を刺激して表に出てくるだけの役割しか果たさないという見方もできるとしている。

そのような観点から HAL と人間との会話を注意深く

観察すると、HAL はもはや理想的で完璧な理解力を持った知能ではなく、単なる「お話マシン」のように見えてくる。実際、複雑な思考を必要とする人間の質問に対して、それを別の方向にいなすよう受け応えする場面がいくつか見られる。その典型例が、自信の欠如に関して問われた場合に完全無欠と答える場面だろう。

単なるお話マシンであるとするれば、HAL 程度の対話をする計算機をつくることはそれほど難しくはないだろう。膨大な対話のデータベースと、その中に納められた個々の対話状況へのインデックス、そしてインデックスの照合によりそれは実現できる。

では、HAL はこのような適当に会話をする計算機としてプログラムされたのか？ それともやはり知能の理想的な完全なモデルとしてプログラムされたのか？

後者の選択には無理がある。なぜなら、物語の後半で HAL は人間に反乱する。船長に命じられたことを拒否するのである。完全な知能を仮定するのであれば、HAL はこの時点までに目標の衝突ということを経験しているはずだ。しかし HAL が今までにそのような状況におかれたという証拠は一つもない。仮にあったとしても、人間の命令にそむくような計算機は、このような重要なミッションには選ばれないだろう。

では、HAL は単なるお話マシンか？ もしそうだとすると、設計者は、この最後の反乱の状況を事前に HAL のデータベースに与えていたことになるが、これは設計者の動機として考えにくい。

以上のことから、Schank は、HAL は完全無欠な知的計算機として非現実的な概念だと結論している。

6. 共同作業

認知科学者 Donald A. Norman は、人間および機械の犯すエラーという視点から機械と人間の共同作業に関して考察している。

個々の技術項目に関する議論の紹介は省くとして、彼の考察で最も注目されるのは、この映画が「テクノロジーに関する底なしの楽天主義」に基づいているという部分である。映画の中に描かれている宇宙での生活には、完全無欠な計算機と日常作業においてほとんど間違いを犯すことがない乗組員が登場する。定期旅客機を操縦する十分に訓練された操縦士であっても、業務中にある頻度で間違いを犯すことが知られているのに、ディスカバリー号の乗組員はそういった間違いはほとんど犯さない。間違いを犯すのは、物語のプロットに重要な場面だけであり、そのことがこの映画の非現実性を感じさせる。

今日、人間と機械の相互作用、そしてそれに関連するさまざまな問題点を克服するための研究が重要視されている。これまでの技術発展を見ても、それらの問題を克服するためのさまざまな技術が開発されてきた。

確かにHALは人間と音声で対話できるが、宇宙船内の他のマンマシンインタフェースを見ると、今日の技術よりもはるかに遅れている部分が多く見られる。それは当時の計算機技術の単純な延長線上にあるような時代遅れのものが多く見られる。ここにおいても、この映画が夢やユーモアに欠如した当時のテクノクラートの知識に支えられてつくられていたことがうかがえる。

テクノロジーを発展させるうえで、技術的要素以上に困難な問題は、人間的・社会的要素である。「2001年」は技術的問題については巧みに処理し表現しているが、この社会的要素を見逃している。巨大で高速な計算機ができれば人間並みかそれ以上の知能が実現できるという考えが背景に見られる。機械と人間の間に適切な相互関係を築くというのは、今日においても非常に困難な研究課題であり、技術的な意味で、「2001年」はそういった観点に基づく描写がほとんどないという特徴を持っているとNormanは述べている。

7. 感情

HALは、人間とのコミュニケーションにおいて、画像理解や音声対話といった機能以外に、二つの特殊な機能を備えているように描かれている。それは感情の表現と認識である。これは、HALが停止される直前に、恐怖を表し必死に懇願するシーンで顕著だ。感情とは何か？ 計算機に感情を持たせるということは、どのような意味があるのだろうか？ そして計算機を情緒的にさせるにはどのような技術があるのか？ これらの点に関してRosalind W. Picardが考察している。

恐怖や怒りのような内的状態は感情の側面であると思われるが、感情の定義に関する総合的な通説は存在していない。感情とは、怒りを伴う心拍数の増加のような生理的変化の体験であるといった解釈や、感情とは認識であって、思考の一つの形態に過ぎないという解釈などが混在する。しかし、感情は他人に伝達可能であり、それは他に共感を生むなど、コミュニケーションにおいてさまざまな効果を与えることに異論はないだろう。

HALは視覚による表情理解や音声の特徴から人間の感情を同定しているようだが、現在、計算機に感情を理解させる研究はどのようになっているのだろうか？

Picardが所属するMITメディアラボでは、顔の表情の認識、情動音声の合成、そして生理的信号からの感情理解など試みがなされている。

例えば、パターン認識を使うことで、限られた被験者の4種類の感情を高い確率で認識できるようになっている。今はリアルタイム処理はできないが、ハードウェアとパターン認識技術が向上すれば、近い将来に即座に認識できるようになるだろうと予測されている。話し方の特徴のいくつかも感情に応じて変化すること

も部分的に解明されてきた。ただ、一つの方法で信頼性の高い感情認識が可能になるとは思われていない。視覚・聴覚といった知覚的な手がかりと、話し相手の行動予測などを含む認識の手がかりを組み合わせるといったことが必要になるだろう。

さて、感情に関する研究が進んだとして、HALのように感情を持つ計算機が可能になった場合、それは何を意味するのか？ 感情的と合理的という概念は対立するもののように思われるが、最近の脳科学の進展によりこれらは密接に関連していることがわかってきた。人間の脳を、新皮質、大脳辺縁系、旧皮質に分ける脳三層構造説と呼ばれる粗いモデルがある。新皮質は高次知能を、そして辺縁系は感情、注意、記憶の中核をつかさどる。辺縁系は、好き嫌いといったヴィランスやまた、セイリアンスの決定に関与し、人間の柔軟性、予測不能性、創造的行動の一因となっている。

新皮質と辺縁系の関係は、例えば支配において固定的ではなく、相互作用を行う。感情には、恐怖や驚きや無意識の反応といった生まれつき持っている一次感情と、二次感情がある。二次感情は、認識にかかわる出来事と低レベルの生理反応を結び付け、新皮質と辺縁系の連合活動の結果として具現化する。このような感情は、合理的判断を含む意思決定に極めて重要な役割をはたすことが知られてきている。

二次感情にとって不可欠な新皮質と辺縁系の間の伝達ルートが損傷などにより失ってしまった人間は特異な行動を見せる。それは、知能指数や認識能力は標準的レベルでも、判断ができなくなる。合理的な可能性を延々と吟味し続けても、適切に合理的に判断し、知的に行動するといった能力を大幅に損なうのである。

つまり、感情（一次感情）が激しいと論理的判断に悪影響が及ぶが、面白いことに、感情（二次感情）が乏しくても論理的判断が損なわれるのである。HALは、基本的に合理的で理論的な判断を行う計算機であるが、上で見てきた人間の場合と同様に、そこには感情が不可欠であったのかもしれない。現在の計算機は、おおむね合理的すぎて、価値判断がうまくできない。

人間のように、バランスのとれた知的な感情のコントロールと他人の感情を認識する能力を、情緒的知能と呼ぶ。この概念は、人間と計算機の相互作用にも応用できる。ただし、相手の感情を理解するだけでは不十分である。情緒的知能とは、感情の認識、表現、所有だけにとどまらず、それらの能力の使い方に関する知識と分別を必要とする。HALはある種の感情を持っているようだが、情緒的知能が低いと想像できる場面がいくつかあることに気が付くであろう。

それから、HALの暴走と感情について興味深い考察がなされている。先に述べたように、HALは、乗組員の誰も知らない真のミッションを事前に知っており、そのことによりジレンマ（ミッションの遂行と乗組員

に嘘をつくこと)に陥ったと見られる。HALといえども計算資源は有限だ。もしHALの資源が、そのジレンマの解消のための手段探索に多く配分されてしまったとしたら、他の業務のために使える資源が少なくなり、エラーの確率が増す可能性がある。つまり、感情がないと適切な判断はできないが、感情がある場合には間違った判断を下す確率を高める。もっとも、HALの暴走はもっと他のことに起因するものかもしれない。

さて、感情は高度な人工知能に有用だとして、果たして一次感情の意味で破壊的な行為にでないような計算機は実現できるのだろうか？ アシモフは、ロボット三原則について述べている。しかし、この三原則は完全ではなく、例えばHALがおかれた状況がそうであったかもしれないように、その三原則に基づいても合理的判断を下せないような矛盾した状況を想像することはできる。状況判断ができ、そして最終的には原則を無視できる感情がなければ、原則に縛られたロボットは動けなくなってしまうだろう。

では、感情を持つ計算機はつくるべきではないのか？ Picardはそうは思っていない。感情を持つことにより適切に判断・行動することと、感情的に振る舞うというジレンマは解決できるだろうという楽観的な予見を、根拠は乏しいが、提示している。Picardは、人間は知的で扱いやすい柔軟な機械を欲しており、さらに感情はそのような機械に不可欠だろうと予測している。ただ、ここで大きな疑問が持ち上がるのだと主張する。それは、感情的計算機を持つにあたって、人間側の準備は整っているのだろうかということである。

8. おわりに

「HAL伝説」に見られる何人かの人工知能研究者の考え方を紹介してきた。「HAL伝説」に収録された論文は、「2001年」に関して、未来予測としての今日の状況に照らし合わせた場合の精度、技術の実現可能性、そして技術内容の必然性・妥当性などに関して議論している。項数の制限や筆者の力不足、そして筆者の趣味などから、すべての論文を紹介することはできず特定の論文に偏ってしまったことをお許し願いたい。本稿で紹介しきれなかった内容としては、フォールトトレランス、画像理解、音声合成、音声認識、チェスプレイ、視話、プランニング、倫理などがある。それらに関しては原著もしくは邦訳を当たっていただければ幸である。

筆者が「2001年」に出会ったのは比較的遅く、1980年代の前半、ちょうど学部を卒業した頃ではなかったかと思う。人工知能という言葉は、それ以前、数学科の学部生だった頃から知ってはいた。筆者のいた理学部には当時既に人工知能の科目があり、興味本位から何度かその授業に出席して人工知能の話聞いていた

からである。授業の中で話された人工知能の定義から、それはどのようなシステムなのか漠然とは理解していたつもりでいたが、実のところ、具体的なイメージが湧かずにいた。そして「2001年」を見て、ひどく感動したおぼえがある。私が漠然とイメージしていた人工知能がまさにそこに描かれているようであった。美しく巧みな映像で構成された映画全体をひんやりとした雰囲気貫き、そこで知能の理想像として設計された人工知能が活躍していた。それが私の「2001年」特にHALに対する第一印象であった。その後、何度もこの映画を見て、また人工知能研究に携わるようになってHALに対する考え方は変わってきているが、初期の頃に大きな影響を受けたことは間違いなし、Blade Runnerと並んで今でも最も好きな映画の一つである。

人工知能の研究に携わる大半の方々は、それぞれ人工知能の将来像を持っていることだろうと思う。逆に、そのようなイメージがなければ、人工知能という領域においてさまざまな研究を集約していくことはかなり困難だと思われる。また、理論や技術の研究過程において、それらが人工知能実現においてどのような意味や価値を持っているのかを将来像に照らし合わせて考えることであろう。先に述べたように、「2001年」の作者は、この物語を未来予測としては位置づけていない。しかし、人工知能研究に携わる少なからずの人が、肯定的、否定的の違いはあれ、この映画に何らかの意味で影響を受けているのではないだろうか。そして、「2001年」は、これからもさまざまな機会に引用されていくと思う。人工知能を扱ったSF映画は多く存在するが、その中でも「2001年」はさまざまな意味で愁眉であり、今日、その内容に関して科学的・技術的な考察するに耐えうる数少ない傑作であるとして、この稿を閉じることにする。

◇ 参考文献 ◇

[Stork 97] David G. Stork, editor. Hal's Legacy — 2001's Computer as Dream and Reality —, The MIT Press, 1997. (邦訳:日暮雅道・監訳,「HAL伝説—2001年コンピュータの夢と現実—」,早川書房,1997)

2000年11月7日 受理

—— 著者紹介 ——



大沢 英一 (正会員)

1959年生まれ。1982年東京工業大学理学部数学科卒業。1982年ソニー(株)。1986～1987年ハーバード大学大学院言語学科。1989年(株)ソニーコンピュータサイエンス研究所。2000年公立はこだて未来大学情報アーキテクチャ学科教授、現在に至る。マルチエージェントシステム、注意と認知、情動理解などの研究に従事。1994年度人工知能学会全国大会優秀論文賞を受賞。日本ソフトウェア科学会、情報処理学会各会員。本誌編集委員。博士(工学)。