

リスク環境における人と知能機械の協調をデザインする

稲垣 敏之

1. はじめに

近年の機械は、状況を認知・解析し、今何をなすべきかを決め、それを実行に移す能力を備えている。このような高い知能と自律性を持つ機械は、日常のいたるところに存在するが、ここでは、交通移動体における自動化システムを例にとり、人と機械の関わりを考察してみることにしよう。

航空機における自動化システムは、当初は簡単な姿勢制御のみを行うものであったが、やがて地上の無線標識施設からの信号を取り込んで精密な操縦を行えるようになる。そして今や制御能力のみならず、機体重量や気象条件などの情報をもとに最適速度・高度を計算し、飛行を計画する能力を備えている。このような自動化システムの進展が航空機の安全性向上に寄与してきたことは明白である。

一方で、高度な知能、自律性、高信頼性といった自動化システムの特質が、人間機械系の安全性を損ない得る問題を惹き起こしていることも、また事実である。

本稿では、人と高度技術システムのミスマッチともいえる問題を示した後、人と機械が自然な形で協調できるシステムを実現するには、システム設計にどのような視点が求められているかを考察する。

2. 人と高度技術システムのミスマッチ

高度自動化の進んだ人間機械系に生じている代表的な問題には、

- (1) 多機能インタフェースによるエラーの誘発
- (2) 人と機械の意図の対立
- (3) 自動化システムへの不信と過信の交錯
- (4) 自動化システムによる「異常の隠蔽」
- (5) オートメーション・サプライズ (automation surprise) などがある⁽¹⁾

(1) は、コンパクトな空間のなかに多くの機能を実現するための「モード」がもたらす不都合である。人は自動化システムに「モード A」のもとでの指示を与えたつもりでいても、実際には「モード B」のもとでの指示であったりすることがある。

(2) は、機械が人の行為の邪魔をするケースである。同じ場面においても、人が見ているものと、機械が見ているものが必ずしも同一ではないことに起因する。

(3) は、警報システムに典型が見られる。警報生成のしきい値を安全側に定めると誤報が増えるが、誤報を何度か経験すると、正しい警報が出ても、「また誤報だろう」と放置することがあり得る。一方、人は、「異常が起これば警報が鳴るはずだ」と、状況監視を警報システムに任せきりにすることもある。

(4) は、自動化システムの特質に起因するところが大きい。自動化システムは、制御対象に生じた異常を覆い隠すほど強い制御をかけることができるが、その制御がどれほど骨の折れるものか、人に伝えようとはしない。したがって、よほど注意しない限り、背後に隠された制御対象の異常に人が気づくことは難しい。

(5) は、自動化システムが知能を活かして「気を利かせる」ことなどによって起こる。あまりに自律的にことが進むと、人は状況が把握できないまま、「いったい何が起きているのだ？」とただ驚くことになる。

3. 人間中心の自動化

人と高度技術システムとの間のミスマッチを回避する手がかりのひとつは、「人間中心の自動化」である⁽²⁾。「自動化できるところはすべて自動化する」、あるいは「自動化するか否かはコスト最小化の観点で決定する」という、技術主導的なアプローチへ

の反省から生まれた「人間中心の自動化」の考え方は、早くから自動化システムが多用されてきた航空分野で議論され、表1のように纏められている。

ただし、「人間中心の自動化」は、まだ十分に精密化されたレベルには到達していない。例えば、表1は、「人に指揮権がなければならぬ」としているが、「いついかなる場合でもそうあるべき」なのか、「ある一定の条件が満たされる場合はその限りではない」のか明確ではない。

また、「自動化システムはオペレータ(人)をモニタできるようになっていなければならない」との主張も、人の行為に不都合が検出されたとき、「警告を発して人に注意を促すにとどめるべき」なのか、「必要に応じて自動的に介入し、人の行為の不都合を是正する措置を講じることを許す」のかは、明らかにしていない。

このように、表1の「人間中心の自動化」には不明確な点が多々あり、解釈次第では、表1の言明間に不整合が生じるだけでなく、人間機械系の設計が大きく異なり、結果的にシステムの安全性も異なることもあり得る。

4. 対象に依存する「人間中心の自動化」

「人間中心の自動化」は一律のものではない。そこで実現すべき「人と機械の協調形態」は、何を対象とするかによって変わり得るものである。例えば、自動車のための「人間中心の自動化」は、航空機のための「人間中心の自動化」と大きく異なっていてよいのだが、そのことは意外に認識されていない。「人間中心の自動化」が対象に依存することは、「運転者の質」と「時間余裕」の2つの視点から考察してみれば明らかである⁽³⁾。

「運転者の質」は、教育・訓練を反映している。定期的に一定の教育・訓練が課される航空機のパイロットと、いったん免許を取得すれば、よほどのことがない限り教育・訓練を課されることがない一般乗用車のドライバの間には大きな違いがある。それは、とっさの判断や操作技量、自動化システムに対する理解度の違い等となって現れる。自動車にも自動化システムが次々と搭載されていく昨今、ヒューマン・インタフェースを含めたシステム設計に十分配慮しなければ、自動車における「人と高度技術システムの不マッチ」は、航空機のそれとは比較にならないものとなる可能性がある。

「時間余裕」とは、緊急時において人が認知・判断・操作に使用できる時間を指す。例えば、他機との衝突を避けるには上昇すべきか降下すべきかを航空機衝突防止装置TCAS (Traffic alert and Collision Avoidance System) がパイロットに提示するのは、脅威機への最接近予想時刻の15~35秒前である。パイロットがそれから5秒以内に操作を開始すれば、衝突は回避できる⁽⁴⁾。しかし、自動車なら、先行車への衝突警報が発せられてから衝突回避操作を始めるまでに許される時間は高々数秒である。すなわち、航空機に関しては、「いついかなる場合でも、人に指揮権を与える」ことが妥当であったとしても、自動車についてもそうであるという保証はない。

5. 知能機械による多層支援

人への支援は多層的に行う必要がある。まず、「平時における支援」としての「状況認識の強化」である。人の認知はつねに正しいという保証はないが、「目の前の状況が的確に認識することができれば、何をすべ

表1 航空における「人間中心の自動化」の原則⁽²⁾

人は、航空システムの安全のため、最終の責任を負う。

それゆえ、

- 人に指揮権がなければならぬ。
 - 指揮を効果的に行うために、人は直接的に関与できなければならない。
 - 人が直接関与するには、人に情報が提供されねばならない。
 - 機能を自動化してよいのは、適切な理由がある場合に限る。
 - 人は自動化システムをモニタできるようになっていなければならない。
 - それゆえ、自動化システムは予測可能でなければならない。
 - 自動化システムはオペレータ(人)をモニタできるようになっていなければならない。
 - システムを構成する各要素は、他の要素の意図に関する知識を持っていなければならない。
 - 自動化は、簡単に学べ、簡単に使えるようにデザインされていなければならない。
-

きかは自ずとわかる」ともいわれることから、「状況認識の強化」は人の意思決定の支援でもある。また、意思決定をするのは人であるという意味において、「状況認識の強化」は、「指揮権を人に与える」という「人間中心の自動化」の考え方に整合的である。

しかし、人を取り巻く環境はつねに平穏であるとは限らない。急変する状況の中で、とっさの判断や操作が求められることがある。後述するように、緊急性が高い状況の中で、今何が起こっているか、何をすれば事故を防ぐことができるかがわかっているにもかかわらず、人はその事故回避操作を実行できるとは限らない。そこで「緊急時における支援」としての操作支援、あるいは「権限委譲の作りこみ」が必要となる。ただし、ここでいう操作支援には、機械の判断に基づいて実行されるものが含まれることから、「権限委譲の作り込み」は、いわゆる「人間中心の自動化」の考え方に整合しないのではないかと議論も起こり得る。本節では、その点についても考察してみよう。

5. 1 状況認識の強化

状況認識の対象は多様であり、外部環境、運転者自身、制御対象（移動体）、自動化システム等が含まれる。第2節に述べた人と高度技術システムの不マッチを避けるうえで、状況認識を強化するヒューマン・インタフェース設計にはつぎのような配慮が必要である。

- (1) 機械がなぜそのように判断したのか、その根拠が分かるような情報の提示。
- (2) 機械が今何をしようとしているのか、意図理解の手がかりとなる情報の提示。
- (3) 人と機械の状況認識共有を助けるような情報の提示。
- (4) 機械への過信を防ぐために、機械の

能力限界を知る手がかりとなる情報の提示。

これらの重要性は、つぎの例から想像できよう。

(例1) 高速道路走行におけるドライバーの負担軽減を図るシステムにアダプティブ・クルーズ・コントロール(Adaptive Cruise Control: ACC)がある。ACCは、ドライバーの希望速度による一定速走行あるいは先行車との車間距離の保持を行う能力を持つ自動化システムである。さて、ACCを使って先行車を追従してきたところ、先行車が減速を始めたとする。前方が渋滞しているらしい。ただ、隣のレーンはまだ順調に流れているのか、高速で走る車が前方に見える。このようなとき、「ACCはいったいどの車を先行車として認識しているのか」を明確に知りたいと思うことがある(図1 Case 1)。もし、追従対象が現在の走行レーン上の先行車であるならば、やがてACCの減速操作が始まるはずである。しかし、隣のレーンの前方にいる車を追従対象として認識しているなら、ドライバーがACCを解除するか、自らブレーキ操作を行う(これによってもACCは解除される)ことをしない限り、走行レーン前方の先行車に異常接近することになる。

また、ACCで先行車を追従中、隣の車線前方からやや速度の遅い車が見えてきたとき、「ACCにはその車が見えているか」気になることもある(図1 Case 2)。割込もうとしている車がACCに見えているなら、その車との車間を確保すべく、ほどなくACCは減速を開始するはずである。しかし、割込み車がACCに見えていないなら、このまま放置すれば、割込み車に急接近することになる。

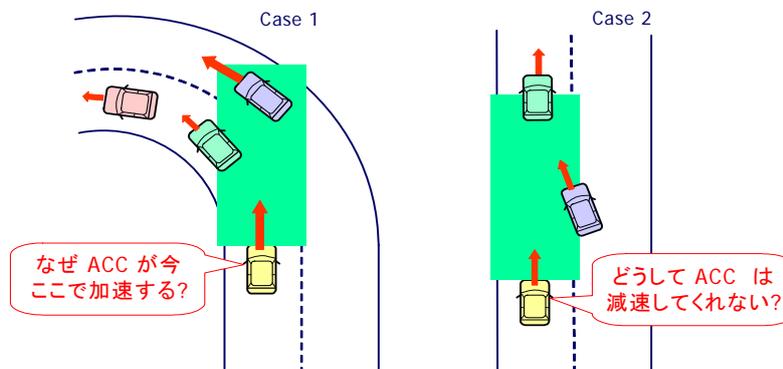


図1 オートメーション・サプライズ

このように、ACCには何が見えて、何が見えないのか、今何を見ているのかがわかれば、ドライバはACCの意図・挙動を予測することができるため、ACCに対して過信や不信（不安）を抱くこともなければ、オートメーション・サプライズに遭遇することもないはずである。

5. 2 権限委譲の作り込み

「状況認識の強化」の目的は、人に状況を的確に理解させることにより、「危険な状況」に立ち至る可能性そのものをあらかじめ排除しようとする点にある。しかし、人が注意を持続させることはむずかしく、ふとした拍子に「注意散漫」と称される状態に陥ることもある。そのようなとき、眼前に潜在的な危険が迫りつつあることを機械が検知できれば、警報などで人の注意を喚起する必要もある。警報の発出も「状況認識の強化」に含まれる。

しかし、それだけで十分だろうか。図2は、自動車の正面衝突ならびに追突事故に関する359件のデータをもとに、ドライバがどのようにして衝突を回避しようとしたかを示したものである⁽⁵⁾。この図は、眼の前に迫りくる危険に気がついていても、すなわち、状況認識は正しくできていても、人は緊急時には必ずしも的確な危険回避操作を実行することができないことを示している。特に、何も操作ができないまま衝突していくケースの多さは、「人には能力の限界がある」ことを直視したシステム設計の必要性を物語っている。

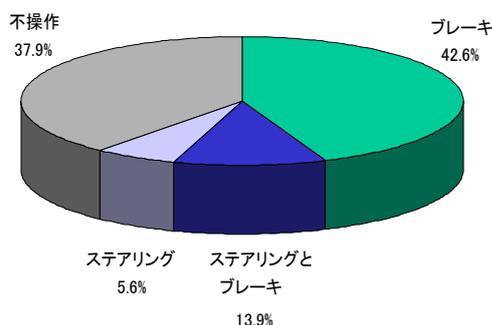


図2 緊急時の危険回避操作⁽⁵⁾

仮に、「緊急時においても最終的な指揮権を人に与える」ようにシステムが設計されていたとしよう。そのとき機械に許されることは、人が事故回避操作を完了するまで、「追突注意!」、「ブレーキをかけよ!」な

どの警報を発し続けることだけである。やかましく警報は発するものの、決して手を差し延べようとしめない傍観者的な機械は、はたして「知能機械」と呼べるだろうか。そして、人は、そのような傍観者的な機械をチームメイトとして認めるだろうか。

そこで必要となるものが「権限委譲」である。権限委譲とは、ある主体（人または機械）が行っていたタスクを、ある時点で別の主体（機械または人）に譲り渡すことをいう⁽⁶⁾。人の能力限界を補う意味での権限委譲は、人が行うべきタスクが適切に遂行されていないことが分かったとき、そのタスクを機械が引継いで遂行する形をとる。

権限委譲は「いつ、誰から誰へ権限を移すか」を問うものであるが、「その判断を下し、権限委譲を実行するのは人か、機械か」の検討も重要である。権限委譲の方式によっては、「指揮権を人間に与える」ことを主張する「人間中心の自動化」の考え方に抵触することがあるからである。しかし、人の特性や能力の限界を考慮すると、「指揮権を持つべきは人か、機械か」という、人と機械を対立させた二者択一的なシステム設計論からの脱却、すなわち、「あるときは人が指揮権を持ち、またあるときは機械が指揮権を持つことによって人の能力の弱点を補完する」といった、状況に即した柔軟なシステム設計が求められる⁽⁷⁾。

5. 3 最適な自動化レベルの選択

そのようなアプローチのひとつが、さまざまな文脈を想定し、そのなかでの人と知能機械の相互作用を記述し、それらが作り出すリスクを定量的に解析・評価することによって最適な人と機械の協調のあり方を探ろうとする方式である。そこで重要な役割を演じるのが「自動化レベル」の概念である。Sheridanによる古くから知られた10段階のリスト⁽⁶⁾に、筆者らの提案⁽⁸⁾による「自動化レベル6.5」を加えたものを表2に示す。

表2は、人と機械の協調形態をシステムティックに考え出す手がかりとして利用できる。そのことを示そう。

(例2) ACCを用いて高速道路を走行中、センサが先行車の急減速(自車ACCの通常ブレーキでは対応できないほどの減速)を検知したとする。この場面に対するドライバ支援として、どのようなものが可能だろうか。

与えられた文脈（先行車の急減速）のもとで、表2の自動化レベルの各々が、どのような「支援形態」を表しているのかを考えてみると、11個の支援形態のうちで現実的なものは、つぎの4つであろう。

- (1) (1)「ブレーキをかけよ」と警報を発する方式。この方式の自動化レベルは4である。
- (2) (2)「ブレーキをかけよ」との警報が発せられてしばらくしても（例えば、2秒経過しても）ドライバがブレーキをかけなければ自動的に緊急ブレーキをかける方式。この方式の自動化レベルは6である。
- (3) (3)「ブレーキをかけます」と宣言すると同時に自動的に緊急ブレーキをかける方式。この方式の自動化レベルは6.5である。
- (4) 自動的に緊急ブレーキをかけた後で、「今、緊急ブレーキをかけました」と事後報告する方式。この方式の自動化レベルは7である。

このうち、権限委譲を事後報告する形をとる(4)は、オートメーション・サプライズを惹き起こす可能性がある。断りなしに、いきなり緊急ブレーキをかけるからである。これに対し、自動化レベル6.5は、自動化レベル7の特質である操作の迅速性を確保しつつ、緊急操作（今のケースでは自動的な緊急ブレーキ操作）を行うことを人に通知することで自動化レベル7の欠点（オートメーション・サプライズ）の解消を図ろうと考案された⁽⁸⁾のものである。すなわち、自動化レベル6.5は自動化レベル7の方策に劣る点はないことから、例2に対する実質的な代替案は、(1) - (3)となる。

では、これらの中から最適なものを選ぶにはどうすればよいだろうか。それには、3つのアプローチがある。第1は、数理モデルを用いた理論解析による方式⁽⁹⁾、第2は、支援システムをドライビングシミュレータやフライトシミュレータなどの上で実現し、被験者を招いて評価実験を行う方式⁽¹⁰⁾、⁽¹¹⁾、第3は、心的特性を考慮に入れた人の行動モデルをコンピュータ上に構築し、それを被験者の代わりに用いてシミュレーション実験を行う方式⁽¹²⁾である。このうち、第3のアプローチは、「必要とする機能を確実に提供してくれた」という支援システムの信頼性(reliability)の実績がやがて人に高い信頼感(trust)を抱かせるが、ときにはそれが高じて、「これからも、必要な機能は確実に提供してくれるはずだ」という過信に至ることもあるといった状況を想定して設計代替案を評価するのに適している。

「人に指揮権を与える」ことにこだわるなら、自動化レベルは5以下に設定しなければならない。上に挙げた「先行車の急減速」に対処する方式のうち、それを満たすものは(1)のみである。しかし、ドライバがACCに過信を抱く可能性があることを考慮したコンピュータ・シミュレーション(第3のアプローチ)による評価実験では、安全性確保の観点からの最適なドライバ支援形態は、自動化レベル6.5の(3)であった⁽¹²⁾。これは、「指揮権を持つべきは人か、機械か」は、状況依存的事であることを示している。同様の結果は、認知工学的実験(第2のアプローチ)⁽¹⁰⁾、⁽¹¹⁾、数理モデルを用いた理論解析(第1のアプローチ)によっても得られている⁽⁹⁾。

表2 自動化レベル⁽⁶⁾、⁽⁸⁾

-
- | | |
|-------|--|
| (1) | コンピュータの支援なしに、すべてを人間が決定・実行。 |
| (2) | コンピュータはすべての選択肢を提示し、人間はそのうちのひとつを選択して実行。 |
| (3) | コンピュータは可能な選択肢をすべて人間に提示するとともに、その中のひとつを選んで提案。それを実行するか否かは人間が決定。 |
| (4) | コンピュータは可能な選択肢の中からひとつを選び、それを人間に提案。それを実行するか否かは人間が決定。 |
| (5) | コンピュータはひとつの案を人間に提示。人間が了承すれば、コンピュータが実行。 |
| (6) | コンピュータはひとつの案を人間に提示。人間が一定時間以内に実行中止を指令しない限り、コンピュータはその案を実行。 |
| (6.5) | コンピュータはひとつの案を人間に提示すると同時に、その案を実行。 |
| (7) | コンピュータがすべてを行い、何を実行したか人間に報告。 |
| (8) | コンピュータがすべてを決定・実行。人間に問われれば、何を実行したか人間に報告。 |
| (9) | コンピュータがすべてを決定・実行。何を実行したか人間に報告するのは、必要性をコンピュータが認めたときのみ。 |
| (10) | コンピュータがすべてを決定し、実行。 |
-

6. む す び

人を支援するシステムが、能力以上に信頼されても、能力を不当に低く評価されても、支援システム導入の企図は達成されない。人と機械の協調関係を円滑で効果的なものにするためには、「状況認識の強化」と「権限委譲の作り込み」が欠かせない。直面する状況を人と機械がどのように認識し、機械自身のどのような判断のもとで、あるいは人の指示のもとで機械が何をしようとしているのかが人に明確に提示されておれば、機械の意図がわからないことによるオートメーション・サプライズの発生を抑制することができる。また、人と機械の意図対立が起こったとしても、人がその原因に気づくことが可能であろうし、人が機械のモードを誤解しても、それに気づくことは難しくないのである。制御対象の異常を隠すほど強力な制御がかけられていたとしても、機械の意図が「通常の意図」と異なることから分かるはずである。さらに、機械の意図に加えて機械の能力が人に分かっているおれば、機械への不信や過信の根本的要因が除去されたことになる。

なお、「権限委譲の作り込み」の検討には、人の心的特性や知能化機械との相互作用に関わる確率的リスクモデルなどが必要となるが、それは、登場以来 20 年近くが経過しているにもかかわらず、未だに「スローガン」的なレベルにとどまっている「人間中心の自動化」の概念の精密化にも寄与するものと思われる。

文 献

- (1) 稲垣敏之, “ヒューマン・マシン・システム—高信頼性が損う安全性,” システム/制御/情報, vol. 41, no. 10, pp. 403-409, 1997.
- (2) C.E. Billings, Aviation automation – The search for a human-centered approach, LEA, 1997.
- (3) T. Inagaki, “Design of human-machine interactions in light of domain-dependence of human-centered automation,” Cognition, Technology & Work, Vol.8, pp. 161-167, (2006).
- (4) 稲垣敏之, “絶対安全への挑戦: 人を支援する知的技術,” 人工知能学会誌, vol. 17, no. 6, pp. 732-737, 2002.
- (5) 交通事故総合分析センター, 交通事故例調査・分析報告書, 2003.
- (6) T.B. Sheridan, Telerobotics, Automation, and Human Supervisory Control. MIT Press, 1992.
- (7) T.Inagaki, “Adaptive automation: Sharing and trading of control,” Handbook of Cognitive Task Design, Chapter 8, LEA, 2003.
- (8) T.Inagaki, et al, “Trust, self-confidence and authority in human-machine systems,” Proc. IFAC HMS, pp. 431-436, 1998.
- (9) T.Inagaki, “Situation-adaptive autonomy for time-critical takeoff decisions,” Int. J. Model. & Simul., vol. 20, no. 2, pp. 175-180, 2000.
- (10) T.Inagaki, et al, “Automation and human interface for takeoff safety,” Proc. 10th Int. Symp. Aviat. Psych., pp. 402-407, 1999.
- (11) N. Moray, et al, “Adaptive automation, trust, and self-confidence in fault management of time-critical tasks,” J. Exp. Psych: Applied, vol. 6, no. 1, pp. 44-58, 2000.
- (12) T.Inagaki, et al, “Human interaction with adaptive automation: Strategies for trading of control under possibility of over-trust and complacency,” Proc. HCII-Augmented Cognition (CD-ROM) 2005.