

2 1. 発話音声で個々人の日常的健全性の確認を

監視通信領域 塩見 格一

1. はじめに

1994年4月26日の午前8時16分過ぎに、名古屋空港に着陸しようとしていた中華航空140便（エアバス式A300-600R型機）がILS進入中に失速・墜落した。原因の一つはエアバス機の操縦システム・デザインが、パイロットのヒューマン・ファクターに整合していなかったことであり、一躍“ヒューマン・ファクター”という言葉が脚光を浴びることとなった。当時、電子航法研究所は、視聴覚的な仮想現実感を利用した飛行場管制シミュレーターの整備を進めており、この施設の機能を利用した最初の仕事が、航空事故調査委員会からの依頼を受けての中華航空140便のILS進入から墜落までの挙動を視覚化する（コンピュータ・グラフィックス技術により、フライト・レコーダに記録された情報からアニメーション・ビデオを作成する）ことであった。

同時に、筆者らは、航空事故調査委員会からは、コックピット・ボイス・レコーダに記録されている音声についても、パイロットの声だけを自動的に取り出す装置の開発の可否を問われ、以降1998年まで、そのようなソフトウェアの実現可能性を調査した。上手い方法が見つからない俁に、1998年5月頃であったように思うが、指尖脈波のカオス生からアルツハイマー病の進行状態を評価しようとする研究が、飛行場管制シミュレーターの整備に係る当所の共同研究相手方の一つオージス総研（株）と福岡の（株）コンピュータ・コンビニエンス社において進められていることを知った。筆者は、オージス総研の広瀬尚三氏に、カオス論的な信号処理技術を音声信号処理に適用してみることを依頼した。

約半年後、広瀬氏から「カオス論的な信号処理パラメータを適正に設定できた場合には、発話者の心身状態の変化に対応するような、（発話音声から算出される）リアプノフ指数の変化が観測される。」旨の報告を受けた。^[1]

2. 発話音声におけるカオス論的な特徴量

心身状態の変化と音声信号から算出されるリアプノフ指数の変化の関係は、単純なものではなく、数時間以上の朗読の継続によってもリアプノフ指数に明確な変化等が認められない被験者も何人もいて、「そもそもリアプノフ指数とは何なんだ？」と言う状況であって、その状況が以降何年も続くことになったのだが、当時は、「発話音声から発話者

の心身状態（覚醒度や、お疲れ度、等々）が評価できれば、革命的な予防安全システムの構築が可能になる。」と、関係者皆んなで舞い上がっていた。

幾つかの新聞の取材を受けて、ロイターやBBCがやって来て“日本の期待される21世紀の技術”として世界中に配信され、タイムラー・クライスラー社、フィアット社、エアバス社、他から多数の問い合わせがあり、米国大使館からは日本語の論文で構わないから可能な限りの資料が欲しい旨の依頼も受けた。皆が、“革命的な予防安全システム”に期待した数ヶ月であったと思う。

しかしながら筆者らは“革命的な(?)”現象を発見した段階であって、これを説明するモデルも持ち合わせていなければ、難しいカオス理論に関する理解も全く不足していた状況であって、残念ながら、具体的な次のステップへ結び付くようなことは何もできなかった。

リアプノフ指数を算出するまでの信号処理手法に問題があって、「リアプノフ指数そのものが音声から心身状態を評価する指数としては適当ではない。」ことが理解されるのは2003年になってのことであった。

時系列信号からリアプノフ指数を算出する演算は、2016年時点では、パソコンで容易にできるかもしれないが、2000年当時には、音声信号を48.0kHzでサンプリングした場合、継続時間が1.0秒間の音声データであっても、4次元の埋込空間にストレンジ・アトラクタを再構成した場合、パラメータの設定に依存するが、アンサンブルの計算に数時間から十数時間を要することは珍しくなく、5次元以上の埋込空間想定することは、途方もない処理性能のコンピュータを擁さない限り、現実問題として不可能であった。

そのような状況において、2000年12月、“革命的な予防安全システム”の可能性にご共感をいただいた三菱スペース・ソフトウェア社とクレイ社（旧テラ・コンピューター社）から、従来のベクトル型やMPP型とは全く異次元の性能を実現したMTA-2システムの提供を受けて、また併せてリアプノフ指数算出アルゴリズムの高速化実装の共同開発を開始することができた。クレイ社の小林達氏やジョン・フェオ氏のご協力により、2002年、リアプノフ指数に代わる指数としてCEM/M（脳活性化指数；Cerebral Exponent micro/Macro）を定義すること

ができて、これを高速に算出するアルゴリズム SiCECA (Shiomi's Cerebral Exponent Calculation Algorithm) を開発することもできた。後に、当時、MTA-2 システムの日本への輸入には米国議会の承認が必要であったと聞いた。



CRAY MTA-2 System

CRAY MTA-2 システムの導入により、48.0kHz でサンプリングした継続時間 1.0 秒間の音声データは約 1 秒で処理できるようになった。^[2]

3. CEm/M の意味する処の調査研究^[3,4]

新たな指数 CEm/M と SiCECA アルゴリズム以前、リアプノフ指数やその時間的な平均値を計算して発話者の心身状態を評価しようとしていた当時、診断値を得るためには 3 分程度の発話音声が必要であって、アンサンブルの 1/10,000 程度の計算に最高性能のパソコンで 5 分程度を要していた。このような状況において、音声収録実験では、被験者さんの音声を収録しては、被験者さんに実験の目的等を説明しながらパソコンの処理結果を待っていた。

CRAY MTA-2 システムの導入により、新たな指数 CEm/M と新たなアルゴリズム SiCECA を実現し、さらに SiCECA における信号処理パラメータの適正化を進めることができた。パソコン用の CPU の高性能化もあって、2005 年、Apple 社のデスクトップ・パソコン Mac Pro により、継続時間 1.0 秒間の 24bits @48.0kHz の音声データは 1 秒以内に処理可能となり、心身状態の評価に必要なデータ・サイズも発話継続時間として 5 秒程度まで短縮することが可能となった。

当所は 2000 年から、音声分析技術の可能性の調査として、東北大学との共同研究を行っており、2003 年からは鉄道総研とも鉄道運転士の“居眠り防止技術”の開発を目的として共同研究を行っていた。2004 年には、筑波大学の稲垣敏之教授をリーダーとした、小泉首相の掲げた 21 世紀の日本を支

える 15 の研究プロジェクト「政府が科学技術により解決すべき 15 の課題」の一つ「状況・意図理解によるリスクの発見と回避」の一項目として、当所は「カオス論的音声信号処理アルゴリズムと実験用音声信号処理システムの開発 (2004-2007)」を開始していた。

2004 年以降、ヒューマン・ファクターに起因し社会的にも深刻な問題として捉えられた事故は、2005 年 4 月 25 日の「JR 福知山線脱線事故」、2007 年 2 月 18 日の「吹田スキーバス事故」、2012 年 4 月 29 日の「関越自動車道高速バス居眠り運転事故」、2016 年 1 月 15 日の「軽井沢スキーバス転落事故」と、決定的な改善策のないままに今日に至っているが、2004 年当時、高速道路での居眠り事故は深刻な社会問題として認識されており、上記のような特別な予算を投じた研究が進められていた。なお、一参加者としての筆者であったが、また（当時においても）この研究内容の全てに納得していた訳ではなかったが、担当部分に関しては、共同研究者であった東北大学の藤本先生や鉄道総研の佐藤研究員と共に最大限に努力・注力し、十分な成果を上げたことと確信している。

上記の共同研究により、「いつ居眠りを起こしても不思議ではないほどの過労状態においては、その者の発話音声から算出される CEM 値は、健常状態と比較して著しく低下する。（健常状態で計測される CEM 値の平均値と標準偏差を a と σ とすれば、過労時の CEM 値は $a - 2\sigma$ 程度以下になる。）」ことが実験的に確認された。

CEM 値の挙動は実に複雑で、2016 年時点においても明確ではない事柄はいくつも存在している。朗読音声のように発話が主作業である場合の発話音声から算出される CEM 値と、鉄道やトラック等運転士の喚呼音声（「出発進行！」など）から算出される CEM 値とは異なる変化パターンを有している。

いすゞ自動車の協力を得て苫小牧で実施した実車によるテストコース走行実験の結果からは、漫然運転により覚醒度が低下した場合に、喚呼音声から算出される CEM 値が低下した運転士もいれば、蛇行等の運転パフォーマンスが低下しても、CEM 値に明確な変化が見られなかった運転士もいた。後者においては、喚呼発話を忘れないことにより注力した結果が現れたのであろうか？

4. 様々なトピック

筆者らは、1998 年音声のカオス論的な現象の発見から 2000 年頃まで、「エンジンやモータが壊れる前にノイズを発生するように、人間も草臥れて来る

とカオス論的ノイズを発生する。」と考えていた。2016年の時点では「エンジンやモータの回転数を上げるとノイズが大きくなるように、人間の頭も回転数が上がると（活性度が上がると）発話音声に含まれるカオス論的なノイズが大きくなる。」と考えている。音声信号の波形を見ればノイズが大きい状態では“ゆらいでいる程度”が大きくなるように見えるため、嘗ては、リアプノフ指数を時系列信号の“ゆらぎ”の程度を示す尺度と見做していたが、“ゆらぎ”の定義を明確に示すことができなかつたために、このような表現では、逆に、上記のモデルの理解を妨げてきたように思われる。

“疲労”についても、当初の理解が必ずしも正しくはなかつたために、「発話音声から発話者の疲労度が直接に計算できる。」との誤解を招き、この誤解については、未だに解消することができていない。

“疲労の計測”を社会が強く欲していることは間違いのない処であるために、「たまったお疲れ測ります！」と言った単純な表現が一人歩きしてしまった。そもそも“人間の疲労”を定義することは容易なことではなく、10年以上に及ぶ研究の結果としては自明に過ぎるが、過労防止システムや、居眠り防止システムを実現しようとする場合、“疲労”の概念によるのではなく、“計測可能なパフォーマンス”を尺度とする必要があることが明らかになった。

2008年には、昔の黒電話程度の大きさの発話音声分析装置を実現した。朗読カードを提示し、これに対する朗読音声を収録し、3~5秒でCEM値を算出した。米国大使館においてデモを実施し、12月にワシントンD.C.で開催されたFAA安全フォーラムに招かれ日本から初参加・初出品した。以降、米国陸海軍合同睡眠研究所に提供している。

また別ルートであったがフランスの事故調査委員会にも提供し、ボルボ社からのリクエストに対応しスウェーデン語版も製作した。^[5]



Desktop CENTE

Desktop CENTEは管制業務官署等への試用設置を目的として製作したものであったが、機能評価予備実験において、信頼性に問題があると判断されたために、試用設置は実現できなかった。予備実験において明らかになった問題は、立て続けて朗読された音声から算出されるCEM値に大きなバラツキがあるということであった。血圧や体温を続けて10回計測したとすれば、最初の数回は兎も角としても、計測の途中で何事もなければ、それ以降は似たような値が計測される筈であるが、CENTEの場合には、健常状態で計測されるCEM値の標準偏差の3倍以上の差異がランダムに発生していた。この問題は、CEM値を複数回計測した場合には必ず発生した問題であり、CEM値算出アルゴリズムに問題があることは明らかであったが、原因を理解し解決するまでに5年を要する難問であった。

5. 脳と“疲労”に係る考察

“疲労”の概念が一般的過ぎるために、考察する人の数だけ定義が存在しそうで、“疲労”の定義からこれを管理する一般的な、また具体的な装置を実現することはほとんど不可能と思われる。しかし“疲労”に係る考察が無駄な訳ではない。

筆者には、いつ頃からのことか分からないが、嘗ては認知心理学のように言われた科学が、その守備範囲を広げて、今日、脳科学と呼ばれる分野が存在するようである。筆者は、脳科学にマスコミ思考な如何わしさを感じるが、自称脳科学者の方々の営業努力の結果、広く認知されてしまったようである。発話音声から算出する指数により脳の機能状態を評価したいとすれば、何等かの脳モデルが必要であり、筆者は、脳のソフトウェア的な構成をモデル化することが必要と考えた。

先ず、脳の機能は目的定義機能と身体制御機能の相互作用として実現されていると考えた。これらの二つの機能は、ハードウェアとしての脳に偏在して存在すると考えられるが、夫々の機能に要する処理容量について個々に制限等が存在する必要はない。逆に、双方の機能は相互にリソースを融通し合っていると考える方が、脳が冗長系であることから合理的である。筆者は、脳機能の自己組織化は、乳幼児期における生存本能による活動の結果として獲得された情報が記憶として蓄積されることにより始まり、言語の獲得による記憶の言語化により目的定義機能として発現すると理解するが、本稿読者諸氏には同意いただけるであろうか。

上記脳機能モデルは、「人が経験を言語化し記憶として蓄積する能力を有することにより人間性を

獲得する。」という事実に整合する。また、個人的な記憶は、言語化されることにより抽象化・普遍化され、人は初めて社会的に考えることが可能となる。

上記脳機能モデルにおいては、人間の“疲労”はこれを経験し、これを言葉により記憶した時から始まると理解される。この脳機能モデルにおいては、肉体的な疲労は「目的定義機能→身体制御機能→身体器官→感覚器/受容器→目的定義機能」の閉ループにおいて目的定義機能野が認識すると理解することができる。精神的な疲労は目的定義機能内部の「記憶と認識の相互作用」や「感覚器/受容器→目的定義機能」の情報処理に係るものであり、目的定義機能野が疲労として認識すると理解される。この脳機能モデルにおいては肉体的な負荷は必然的に精神的な負荷を伴い、また脳の目的定義機能野と身体制御機能野の間に明確な境界が存在する訳ではないから、過大な精神的な疲労により身体制御機能に影響が及ぶ時、精神的な疲労は肉体的な疲労の誘因となる。この脳機能モデルにおいて、「適度な運動が精神活動を活性化させる。」ということは、軽い肉体的な負荷が、目的定義機能野をアイドリングさせるに適切な精神的な負荷を伴っているからであり、また「運動選手が精神的な問題を抱えていると良い成績が出せなくなったり、音楽を聞きながら作業すれば効率的であったりする。」といった現象も、この脳機能モデルは合理的な説明を与える。

人間の脳の目的定義機能は後天的に構築された機能であり、本能の様な歯止を有していない。従って、自らの身体に対してさえ、過労死する程の無理な制御がなされる可能性は常に存在する。不注意や緊張感の喪失を意識的に防ぐことは不可能であり、ヒューマン・エラーは意識することにより防げる性質のものではあり得ない。従って、経常的に当人の意識し得ない脳機能状態を計測し、その健全性を監視する技術こそが、最上流の予防安全技術となる。

6. おわりに

音声信号処理ソフトウェアの開発において、コックピット・ボイス・レコーダーに記録されている音声データから人間の声だけを取り出すソフトウェアの開発は未だ上手くは行っていない。人間の声の成分が含まれる確率の高い部分の抽出等は一部可能になっている。

1998年に筆者らが本研究の契機となった現象を発見に至る調査研究のキックオフ・ミーティングの場に、全くの偶然ではあったが、筆者に中華航空140便の挙動アニメーションを依頼してくれた元航空事故調査官がいらして、筆者らの打合せ内容を聞

いておられたのだが、5年程後に再度お目にかかった時に、「あんなことが本当にできるのか？」と思った、とお聞きした。確かに、1998年当時においては未だカオス論的な信号処理技術は未知の領域であったのだ。

2015年12月、改正労働安全衛生法により、従業員数50人以上の職場環境においてストレス管理が義務化された。2014年度の後半以来、当該義務化による需要を目指して、様々な脈拍や音声からストレス状態を評価するウェブ・アプリやスマートフォン・アプリが多数提供されてきた。現状、これらは玉石混交と表現されるレベルのようであるが、当所の発話音声分析技術も、遂に音声品位の劣化に対するロバストネスを実現し、現在、幾つかの形態において試用可能な状況となっている。

スマートフォンをプラットフォームとした場合、スマートフォンの機種に強く依存する場合があります。標準マイクロフォン（Earthworks #QTC-50）を使用した場合のようなCEM値の分布における安定性は実現されていないが、現状までの利用者からは、「十分に静かな環境で、経常的に計測していれば、CEM値の相対的な変化が、自分の主観的な“お疲れ感”と整合していると感じるようになった。」とのコメントももらえるようになっている。

謝 辞

発話音声分析技術の普及のために、スマートフォンをプラットフォームとするアプリケーションの開発にご協力いただいている三菱スペース・ソフトウェア（株）、また（株）国際テクノロジーセンターに感謝致します。

文 献

- [1] 塩見，廣瀬： 音声から眠気や疲労を検出する試みについて，第37回飛行機シンポジウム，1999.
- [2] 塩見： 発話分析から考える脳機能モデル”感性工学研究論文集，Vol. 4, No. 1, Feb. 2004.
- [3] 塩見，他： 発話音声による疲労状態評価検証実験の手法と結果，日本人間工学会第35回関東支部大会，Oct. 2005.
- [4] 塩見，他： 発話音声によるトラック運転手の心身状態評価手法と結果，日本人間工学会第36回関東支部大会，Dec. 2005.
- [5] R Ruiz, PP de Hugues and C Legros: Analyse de la Voix d'un Conducteur Automobile: Effets Acoustiques de la Fatigue, 10ème Congrès Français d'Acoustique, Lyon, 12-16 Avril 2010.