

# リアルタイム疲労評価システム

管制システム部 塩見格一  
航空システム部 板野 賢

## 1. はじめに

1998年に発話音声のカオス性から発話者の疲労状態を評価することの可能性を発見して以来、発話音声評価システムの試作等、以下のような研究開発を進めてきた。

- (1) 発話音声評価アルゴリズムの開発
- (2) アルゴリズムの高速化とパラメータの最適化に関する研究
- (3) 発話音声評価システムの時間的分解能の向上と高信頼化に対する研究
- (4) 周辺雑音の除去に関する研究
- (5) 発話音声の評価原理に関する研究
- (6) 疲労と脳機能モデルに関する研究

上記のような研究開発の結果、1998年当時は30秒間の発話音声を数分掛けて処理し、その処理結果から発話者の疲労度を判定することの可能性を示すことが精々であったシステムは、今日、リアルタイムに発話者の発話内容に対するストレスを評価することが可能なレベルに進化した。

以下、現状のシステム実現までの研究開発内容と共に、今後の計画について述べる。

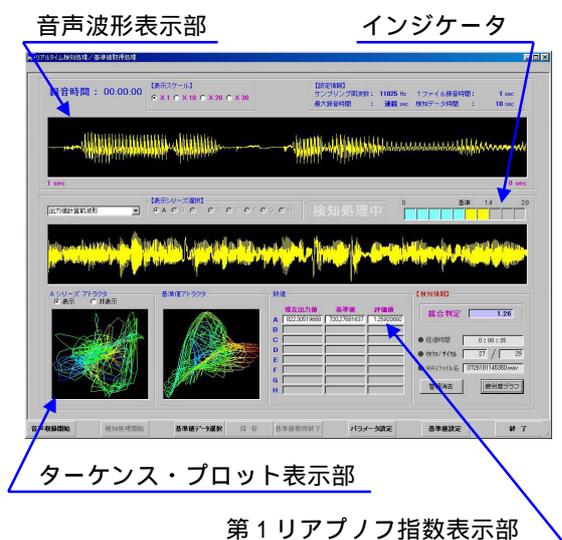


図1 CENTEのGUI

## 2. CENTE

CENTEは当所が(株)オーグス総研との共同研究により試作した、最初のリアルタイム発話音声評価システムである(図1参照)。

### 2.1. 朗読音声の分析

CENTEでは、連続的な発話音声を1秒間の処理単位に分割し、それぞれの処理単位から佐野澤田のアルゴリズムによる第1リアプノフ指数を算出した。この算出された毎秒の第1リアプノフ指数に対して、平均時間幅を60~300秒、また移動間隔を1秒とした移動平均処理を行い、この時間的な平均値を以って発話者の疲労度を評価した。

図2は朗読音声をCENTEにより処理した結果であり、(1)緊張緩和から、(2)疲労蓄積、(3)集中力喪失、(4)疲労中断までの心身状態の変化の様子をグラフ上に見ることができる。

### 2.2. 時間分解能と信頼性

図3はNHK国会中継から首相の所信表明演説を録音し、同様に分析した結果であり、連続的な録音音声を1秒間の処理単位に分割するタイミングを0.1, 0.2, 0.3, 0.4秒とずらした場合の処理結果を重ねてプロットしたものである。

図3からは、第1リアプノフ指数の時間的な平均

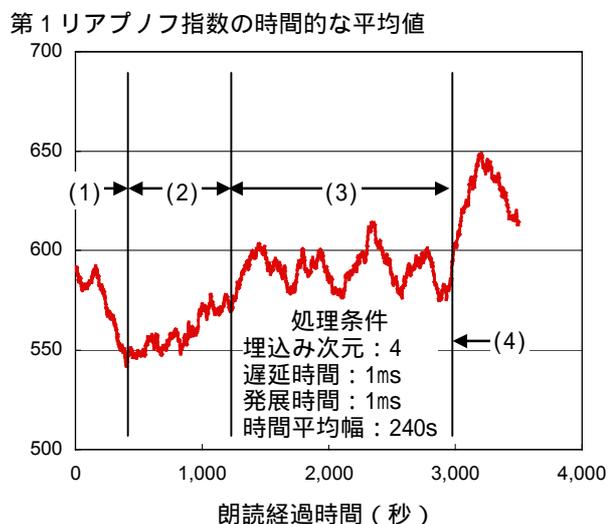


図2 CENTEによる処理結果

### 第 1 リアプノフ指数の時間的な平均値

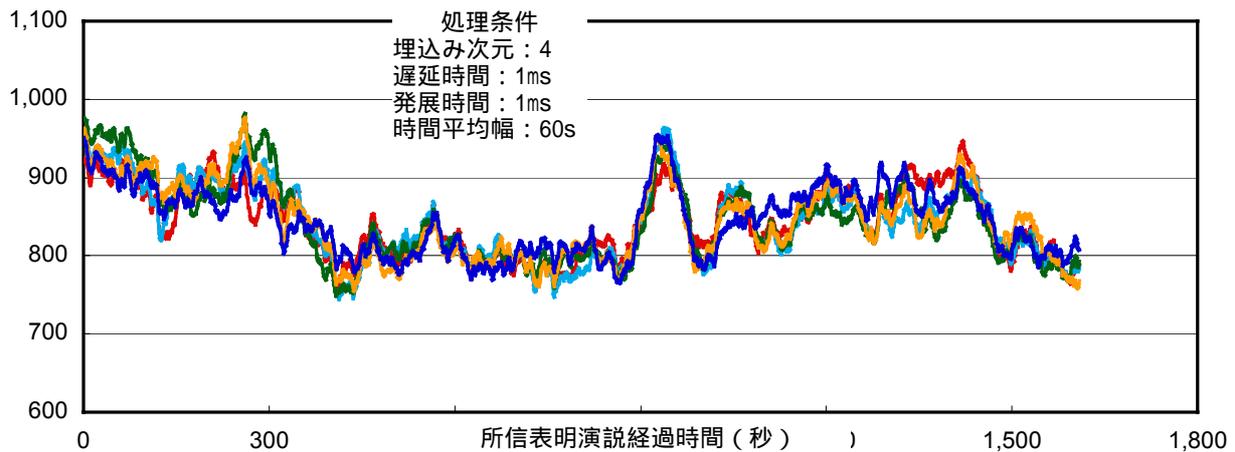


図 3 CENTE における処理単位の切出しタイミングの差異と処理結果

値の変化については、演説開始後 900 秒付近に見られる山や、また演説開始直後から次第に低下する構造は、処理単位の分割タイミングに依らずに保存されており、これらが発話内容に対する意味を有する可能性を伺わせるが、多くのより小さな構造については、「発話内容に対応するような意味を有するとは考えるべきではない」ことが明らかになった。

CENTE による処理では、発話者の疲労状況を評価する程度の目的には十分対応可能であっても、発話内容に対するストレス等、発話のその都度のより詳細な心身状態に係る情報を取出すことはできそうもないと考えられる。

### 3. NONO

NONO は、CENTE の後継システムとして、三菱スペース・ソフトウェア (株) との共同研究により製作したものである (図 4 参照)。

#### 3.1. 信頼性向上のための前処理

録音音声进行分析する場合、CENTE においてはレコーダの再生ボタンを押すタイミングが微妙にずれただけで処理結果に数十パーセントの誤差が生じたため、この問題を解決するために、NONO においては、連続的な音声からの処理単位の切出しを再現性良く行うための前処理プロセスを設けた。

このプロセスは、音声認識技術を利用したものであり、連続音声より予め設定する母音を検出し、その母音の開

始時刻から終了時刻までを処理単位として切出す機能を有している。

この前処理プロセスの実現により、NONO は CENTE に比較して遥かに安定な処理結果を算出するものとなった。しかしながら、汎用 PC の処理性能と佐野澤田のアルゴリズムに依っては、切出される全ての処理単位をリアルタイムに処理することは不可能であったため、NONO では、ハードウェアの処理性能に合わせて間引いた母音の第 1 リアプノフ指数を算出している。

なお、バッチ処理として録音音声进行处理する場合、NONO では音声認識プロセスの識別する全ての母音を処理することが可能であるが、この場合には、音声のサンプリング周波数にも依存するが、リアルタイムに比較して 1 桁以上長い時間が必要な場合もある。

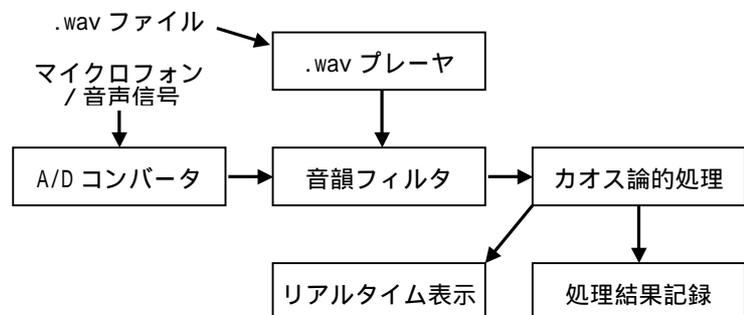


図 4 NONO のアーキテクチャ

### 3.2. システム・パフォーマンス

音声認識技術を利用した前処理は、入力される音声の品位が高い場合には、音声処理の信頼性の向上に有効と思われる処理結果を与えたが、一般的に音声認識システムはかなりの CPU パワーを要するため、第 1 リアプノフ指数計算のためのハードウェア・リソースが相対的に低下し、音声処理システム全体としてのパフォーマンスの向上については、正否の判断は難しい。

特に、高性能と言われる音声認識システムは、高度な辞書システムによりその高性能を実現しており、その結果、無意味な音声を無視したりするので、発話音声の特定の部分のみから第 1 リアプノフ指数を算出している状況も発生する。

音声認識技術を利用した前処理プロセスについては、それが有効である可能性は十分に認められるが、実用的なものとするには解決すべき課題は未だ多く残されていると思われる。

### 4. SiCECA

SiCECA は、佐野澤田のアルゴリズムに代わるものとして、バッチ処理の信頼性を向上させるために開発したアルゴリズムである。

CENTE と NONO はバッチ処理機能も有しているがリアルタイムな使用目的への対応を第一に考えて試作したシステムであり、従来、高速な処理が可能と考えられていた佐野澤田のアルゴリズムを利用しているが、1GHz 程度のクロック速度を有する CPU を使用しても、CENTE では毎秒 1 回、NONO の場合でも毎秒数十回程度しか処理単位を処理することはできない。

発話音声を 44.1kHz でサンプルした場合、連続音声から 1 秒間の音声信号を切出す切出し方は 44.1k 通りであり、このことが如何程の意味を有するかは必ずしも明確ではないが、最も完全なアンサンプルを計算するためには、起点を 1 サンプルずつずらして作成される全ての処理単位に対して第 1 リアプノフ指数を計算しなければならず、完全なアンサンプルを CENTE 或いは NONO で計算しようとするれば、1 秒間の音声の処理に数十分から数時間以上の時間を要する。

発話者の疲労度の評価を目的とする場合には、5 分間の音声から平均的な第 1 リアプノフ指数を求

めることとすれば CENTE で十分であり、数十秒から 1 分程度の時間的な分解能を必要とする場合においても NONO で十分であるが、発話内容を構成する各フレーズに対する発話者のストレス等を評価しようとする場合には、即ち最大限の感度を実現しようとするれば、完全なアンサンプルを求めることが必要となる。(尤も、発話者の発話時のその都度のストレスを評価するためには、完全なアンサンプルを求めるだけでは不十分であり、現時点においては、その十分条件等は明らかではない。)

SiCECA は、発話音声から計算される第 1 リアプノフ指数について完全なアンサンプルを計算する、即ち、サンプルされる音声から切出すことの可能な全ての処理単位に対する第 1 リアプノフ指数を“従来手法によるよりも遥かに高速に”計算するアルゴリズムである。SiCECA は、音声信号が強い周期性を有していることを利用して、第 1 リアプノフ指数算出のための近傍点集合を生成することにより、単純な距離的な条件による近傍点集合探索生成プロセスを必要とする佐野澤田のアルゴリズムに比較して遥かに高速な、処理単位の設定によって 1 桁～3 桁程度高速な、処理を実現することができる。

もっとも、第 1 リアプノフ指数は、不変なダイナミックスがカオス論的な挙動を示す場合に、そのダイナミックスに対して、そのダイナミックスが生成する時系列情報から収束計算により計算されるものであり、連続的な発話音声のように、音声信号を発するダイナミックスが安定ではない場合についても定義されている指標値ではない。従って、厳密な意味では SiCECA により算出される指数値は、第 1 リアプノフ指数ではなく、形式的に計算された第 1 リアプノフ指数に類する指数値であり、我々は、これを“脳機能指数”と呼ぶこととしたい。

物理的な意味においては、脳機能指数は、カオス論的な時系列信号に含まれる、そのカオス性を擾乱させるノイズの強度を示す指標値である。

### 5. Parole

Parole は当所とメディカル・パレットの共同研究により製作した、SiCECA を利要した音声分析ソフトウェアであり、CENTE や NONO とは異なり、録音音声の精密な分析を目的としたソフトウェアである。

### 5.1. 対話音声の分析

図5は、先の演説音声と同様に、NHKの国会中継参考人質疑において、野党議員が与党議員の疑惑を追求している場合の音声を録音し処理したものである。Paroleには、発話者を識別する機能がないので、録音音声に手作業で発話者識別のマーキングを行い処理している。

図5からは、先ず、質問者が応答者以上に、応答者からの言質を得ようと緊張している様子が現れていると考えられる。更に、予め質問者が幾つもの質問項目を用意しており、これを順番にこなして行ったであろうことが、一つの質問を総括して次の質問に移る時に見られるグラフの鋭い山から、明確に読み取ることができる。

また、応答者においては、それぞれの応答において、慎重に言葉を選んでいるためにか、殆どの場合、話し始めに比較して、その終わりには脳機能指数が増大していることが観測され、質問者からの追求により次第に緊張度を増している様子を伺うことができる。

### 5.2. Parole/SiCECAの特徴

Parole/SiCECAでは、処理単位の起点としての処理の可能な全てのサンプルから脳機能指数を算出しているため、即ちサンプルデータに対する完全なアンサンブルを計算するため、その処理結果において、連続音声から処理単位を切出す時刻に起因する差異は原理的に発生しない。

実際には、毎サンプルの脳機能指数の変化をプロットしても、人間の目ではその特徴を把握することが困難であるため、数秒から数十秒程度の時間幅を設定した移動平均処理を行った結果をプロットと

して視覚化しなければならないが、結果的に、Parole/SiCECAにおいては、脳機能指数の時間的移動平均値のグラフは、平均化時間幅と移動時間間隔が同じならば、常に、殆ど同じものとすることができる。

上記において、SiCECAは、先の「毎秒の平均的な第1リアプノフ指数から平均的な疲労度等を評価した」処理方式に比較して、遥かに安定で感度の高い処理結果を実現したアルゴリズムであると言える。

### 6. おわりに

以上、発話音声をカオス論的な手法により分析し発話者の疲労状態や発話内容に対するストレスを評価するシステムの開発について、その経緯と特徴的な成果について紹介した。

分析の対象を発話音声に限定すれば、現時点においてSiCECAは従来いずれの処理手法に比較しても優れたものと考えられ、現在、リアルタイムなシステムへの実装を進めている。

現在、1GHz以上のクロック速度を有するパソコンを利用する場合、44.1kHzで音声をサンプリングした場合、1/64のアンサンブルを算出することができる程度までは開発は進んでおり、またCRAY-MTA型コンピュータを利要した場合には、リアルタイムな処理が可能となる見込みも得ている。

今日、新幹線運転手の居眠り等が問題となっており、将来的には、更に多くの職務分野において業務員のパフォーマンス管理は一層重要な問題になると考えられる。我々の研究開発の成果が、これらの目的に活用されることを強く希望する。

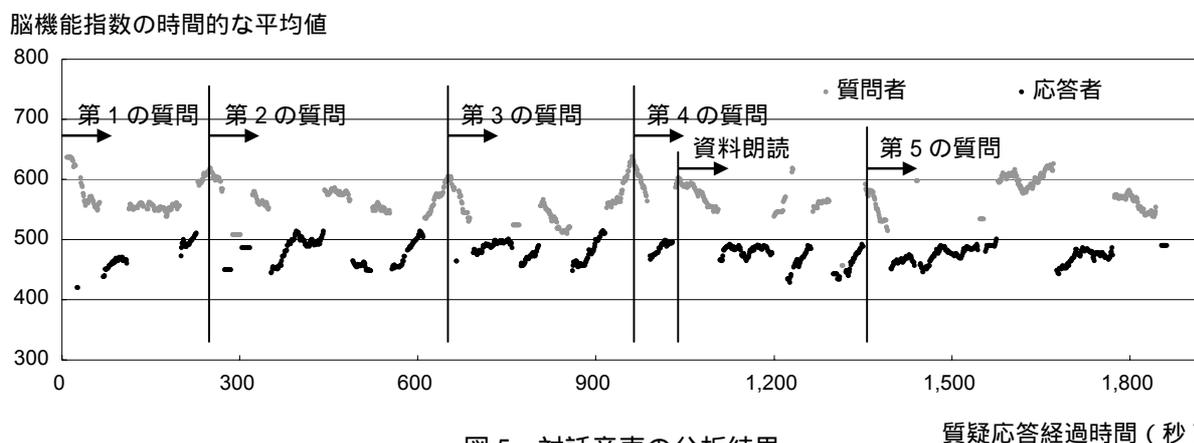


図5 対話音声の分析結果