

12. 発話音声から算出する脳活性度指数の信頼性

機上等技術領域 *塩見 格一

1. はじめに

1998 年に発話音声の“ゆらぎ”の状態が発話者の心身状態に依存し変化することを発見して以来、当所では予防安全装置の実現を目指して発話音声分析装置の研究開発を進めてきた。^[1] その成果として 2008 年には、共同研究先により、発話音声分析装置は「CENTE」という名称の実験計測用装置として製品化された（図 1 参照）。

CENTE の製品化により、比較的容易に発話音声収録実験を実施できる様になり、当所以外の利用者による実験も含めて、より多くの音声データが集まる状況に至っている。その結果、実験環境等に依存し、CENTE の算出する診断値（CEM: Cerebral Exponent Macro 値）に差異とも考えられる傾向が発生する事が明らかとなった。

人間を計測する実験においては、どの様な計測であっても、得られた計測値に高い信頼性を実現することは常に困難であって、その値や値の変化の解釈には高い経験値が求められる。その様な状況において、過去に蓄積データを有さない CENTE においては、その診断値の信頼性・信頼度を示すことは、我々開発者の責務と考えられる。本稿においては、現状までの実験等により明らかになっている CEM 値の信頼性について述べる。



図 1 CENTE の概観

基本的な操作は全て CENTE に実装されているタッチパネルにより可能で、キーボードとマウスは作業精度評価作業の実施等に使用する。

2. CENTE 診断値 / CEM 値

CENTE は図 2 に示す機能構成で実現されており、入力音声の処理経路として赤色の矢印で結ばれる構成要素が診断値としての CEM 値の算出における誤差を生じさせる。マイクロフォンと AD 変換器はハードウェアの個体差としての誤差を生じさせ、デジタル・フィルタと SiCECA (Shiomi's Cerebral

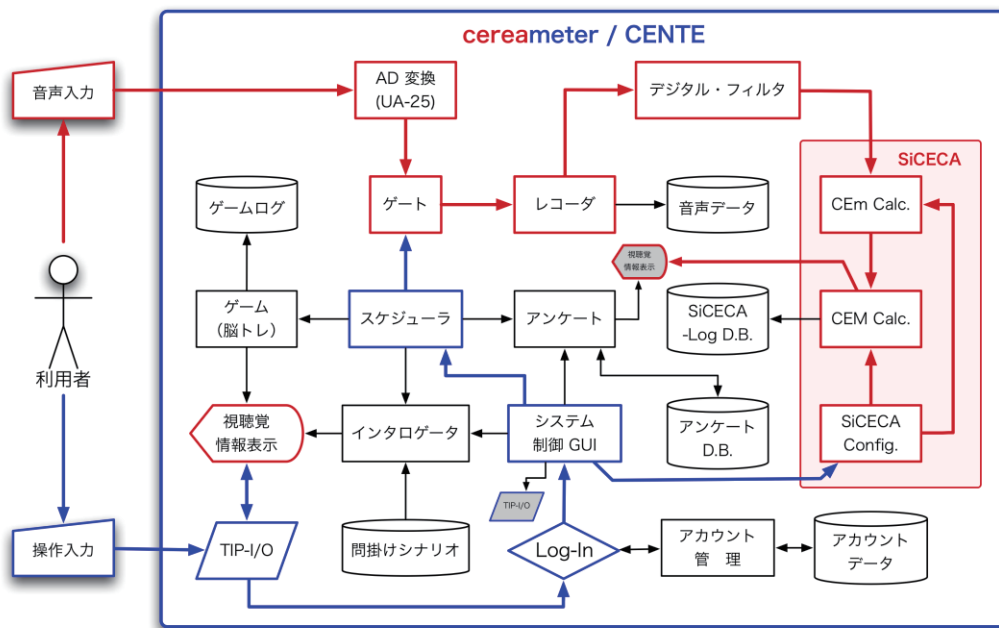


図 2 CENTE の機能構成

Exponent Calculation Algorithm) 信号処理部はパラメータの設定や、CPUの演算処理精度に起因する誤差を生じさせる。

2.1. マイクロフォンの個体差

上記においては、マイクロフォンの個体差に起因する誤差が最も深刻なものであり、周波数特性が極めて良く似ている場合であっても、同一の音声を同時録音した場合に算出される CEM 値が十数パーセント以上も異なる場合も珍しくはない。厳密には2つのマイクロフォンを全く同じ位置に設置することができないので、エコーの僅かな位相差等が影響を及ぼしているのかも知れないが、多少設置位置が異なっても比較的良く似た CEM 値が算出されるマイクロフォンの組合せもあり、マイクロフォンのどの性能仕様項目が CEM 値に強い影響を及ぼしているのか、現時点においては、未だ明らかではない。

現状の技術水準においては、位相特性を揃えたマイクロフォンを入手することは容易ではいが、米国のマイクロフォン製造業者において、電気のスパークを音源としてインパルス特性を揃えたマイクロフォンを用意してもらっており、平成22年度には、これらのマイクロフォンによる CEM 値の計測と比較に関する実験を実施したいと考えている。

マイクロフォンの位相特性を調整することで CEM 値の信頼性の向上が図れることが明らかになれば、そのためのデジタル・フィルタの実現手法や、これに合わせたマイクロフォンの校正環境の実現が可能となるかも知れない。

2.2. 入力音声の問題

発話音声による CEM 算出実験において、スピーカによる再生音声を利用した場合、スピーカとマイクロフォンの設置条件を固定し、同じスピーカと同じマイクロフォンで、同じプロセッサとアンプを使用して、全く同じ音声データから音声を再生した場合に、同じデータレコーダにより音声を収録した場合であっても、収録された音声から算出される CEM 値が同じ値になる訳ではなく、標準偏差として数パーセント（～7%）程度のバラツキが観測される。室温や湿度の差異による影響であるのか、或は否か、現時点では、未だ確認できていない。

今日までの CEM 値算出実験では、上述の様に、再生音声によっても再現性の高い CEM 値の算出が可能である訳ではなかったため、また再生音において生の人間の発話音声に対して何等かの欠落が発生していることを畏れたため、常に生の人間の発話音

声を利用してきた。具体的には、マルチトラック・レコーダ（Sound Device 788T、等々）に複数（8～32）のマイクロフォンを接続し、これに直接に人間が発話した音声を入力することで行って来た。

人間の発話は毎回異なり、算出される CEM 値も毎回異なるため、マイクロフォンの校正は、十秒程度で音読できる内容を記述した朗読カードを利用して、数千回から1万回程度の発話を収録し、個々のマイクロフォンの平均的な CEM 値と、その標準偏差を求めて、これらを相互に比較することで行って来た。現状までの収録データの分析からは、1千回の収録音声から算出される平均的な CEM 値の高い物から順番にマイクロフォンを並べた場合の順番と、2千回、3千回、・・・1万回の収録音声から算出される平均的な CEM 値の高い順番に並べた順番に、8本のマイクロフォンを比較した場合であっても、32本のマイクロフォンを比較した場合であっても、明確な差異は観測されていない。殆ど全てのマイクロフォンにおいて、CEM 値の標準偏差は、CEM 値の平均値の6～8%程度であった。

上記の実験結果から、生の音声データによりマイクロフォンの校正を行った場合の CEM 値の標準偏差と、先の再生音声から算出した CEM 値の標準偏差（7%程度）が、値としては良く似ていることが分かる。必ずしも、これらの誤差が全く同じ原因により発生しているとは言えないかも知れないが、決定的な誤差の低減策が存在しない現状において、生の音声によらなくとも、再生音声により生の音声と同等なマイクロフォンの校正が可能であれば、将来的には、再生音声によりマイクロフォンの校正を容易に、出来るだけ少ない再生音声により、実施できる様に校正環境を整備したい、と考えている。

2.3. 入力音声の信号雑音比やエコー等

上記以外にも、CENTE の設置環境等に起因する誤差も存在し、周辺雑音が大きく信号雑音比が悪い場合や、設置環境がライブでありエコー・レベルが高い場合にも、CEM 値の信頼性は低下する。例えば、図1においては、マイクロフォンやCENTE本体の背面にはフェルト製の吸音パネルを設置しているが、テーブル面からの反射の影響は残っており、マイクロフォンの設置位置がテーブル面に近い（～10cm）場合には、算出される CEM 値が数パーセント以上も低くなることもある。また例えば、周辺雑音が大きき場合には、算出される CEM 値は比較的の高い値になることが確認される。

発話音声を直接に入力する代わりに録音音声を

処理する場合には、非圧縮音声に対して mp2/mp3 圧縮した様な音声では、信号雑音比が比較的に良好な場合であっても、算出される CEM 値は数パーセント（～4%）程度小さな値となる。

2.4. 信号処理ソフトウェアの問題

SiCECA による信号処理プロセスは図3に示す様に、“CEm Calc.”と呼ぶ前半部分と“CEM Calc.”と呼ぶ後半部分から構成されている。

“CEm Calc.”は、変形した佐野・澤田のアルゴリズム（SiCECA アルゴリズムによる前半処理）を全ての音声信号サンプルに時間局所的に適用し、各音声信号サンプルに対して CEm（CE micro）を算出する部分であり、その処理においては主要なものだけでも8つのパラメータ（埋込み次元、埋込み遅延時間、発展時間、・・・）を設定しなければならない。

“CEM Calc.”は、多数の CEm 値に対する統計的な処理（SiCECA アルゴリズムによる後半処理）プロセスであり、非線形の最小二乗法を含み、その処理においては、収束計算の開始近似値の設定等に関して、主要なものだけでも4つのパラメータ（SiCECA- ϵ 、S/N 閾値、・・・）を設定しなければならない。

以上のパラメータは、CEM 値の信頼性や感度に対して、個々が独立な訳ではなく、寧ろ相互に複雑に関連しており、現時点の CEM 値に対する我々の理解においては、最適な組合せを見つけること等は殆ど不可能である。どの様なパラメータの組合せを採用しても、形式的に CEM 値を算出することは可能で、算出される CEM 値だけからは、パラメータの最適化

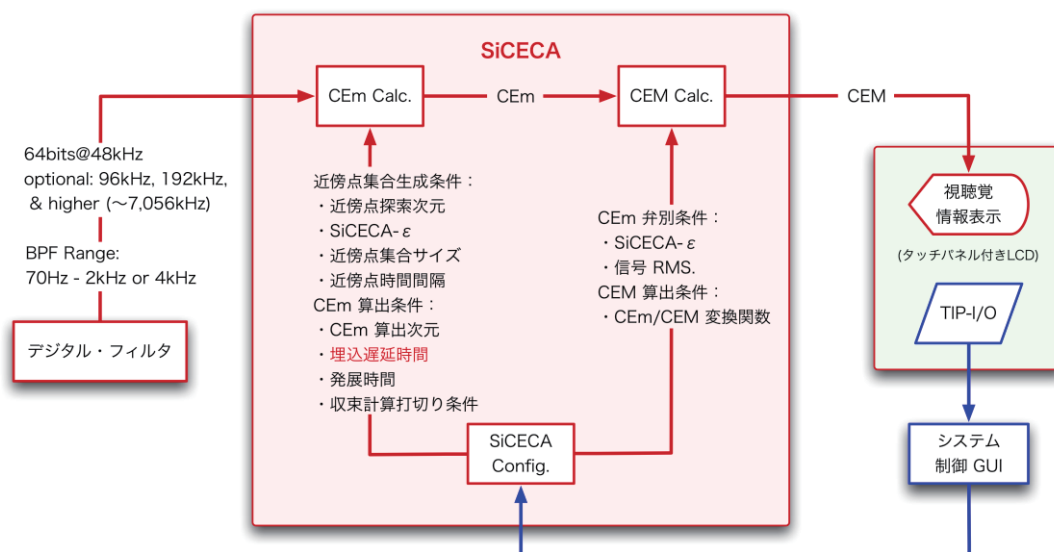
を行うことはできない。

我々は、確実に人間が疲労している状況を生成して、その環境における被験者音声の収録を目的として発話音声により疲労評価実験を実施した。^[2,3,4]

上記疲労評価実験においては、人間の覚醒度に強い相関関係を示すと考えられている臨界フリッカ周波数（CFF 値）を橋本式フリッカ・テスターにより計測し、SiCECA アルゴリズムにおいて幾つかのパラメータの組合せにより CEM 値を算出し、CFF 値の変化と CEM 値の変化の相関が高くなる様にパラメータの調整を行った。もっとも、CFF 値の信頼性も限定的であるため、データ数 40～50 の CFF 値と CEM 値の変化において 0.6 程度の相互相関値が得られる様な、SiCECA 信号処理におけるパラメータの組合せを探した。

上記パラメータの探索作業において発見した最も重要なことは、**パラメータの連続的な変化が、CFF 値と CEM 値の変化の相互相関係数の連続的な変化を与えない**ということであり、我々に取っては、発話音声のカオス性を表象する現象であった。

1998 年当時、我々は実に幸運にも、偶然に、発話者の心身状態に比較的に強く相関する（上記の相互相関係数値として 0.5 程度の）診断値を与えるパラメータの組合せから研究開発を始めることができていた訳であった。埋込み遅延時間（0.02～208.33 ms）と発展時間（0.02～208.33 ms）の2次元空間（埋込み次元は 4、その他のパラメータについては共同研究先との契約により非開示とする。）



1. 埋込遅延時間については、個々の CEm の算出において自己相関係数から設定することも可能とする。
2. CEm は、全ての入力データを時間的な起点として算出することを可能とし、信頼性を考慮しながら間引くことも可能とする。

図3 SiCECA 信号処理



図 4 CFF 値と CEM 値の相互相関係数（部分）

この等高線図において横軸は埋込み遅延時間であり、縦軸は発展遅延時である。この等高線図は、一人の被験者データより、上記パラメータに加えて埋込み遅延次元、SiCECA- ϵ 、等々のパラメータを設定し作成した。

色の濃い部分が、CFF 値の変化と CEM 値の変化の相互相関係数値が 0.4 以上の部分である。

において、見た目では凡そ 8~9 割の部分において相互相関係数値は 0.4 以下であり、特に疲労の検出を目的として発話音声収録していた訳ではなかった状況においては、本当に幸運であった。

3. その他の CEM 値誤差要因

現時点において、上記に加えて更に、単純な音韻の発話においては、母音の違いにより算出される CEM 値は異なり、子音を伴う場合には、子音の差異も CEM 値の差異の原因となることが確認されている。また、朗読音声から算出される CEM 値は、朗読内容による差異を含む可能性があり、想起・換呼音声から算出される CEM 値は、その想起や換呼に係る作業の負荷状態に依存する。

CEM 値には、性差が存在することもほぼ明らかであり、年齢による差異の存在の可能性も無視はできない。

母音や子音の差異、また朗読カードにおける音韻構成の差異については、SiCECA 信号処理パラメータの設定に依存して、それらの差異の大きさや CEM 値の大小関係が変化すると思われる状況が確認されている。CEM 値の算出は、発話者の覚醒度等の心身状態の定量化にある訳で、その求める機能に対する診断感度を維持しながら、音韻構成等に起因する

CEM 値の差異を低減解消することが可能であれば、実用的には好ましい一步の前進となる。しかしながら、現実問題として診断感度の評価は容易ではなく、我々は、差異の解消を目的としたパラメータの調整が、本来第一に求めるべき診断感度の低下に終わらないように注意しなければならない。

このことは、性差等に係る対応においても同様である。

4. おわりに

筆者は、例えば、一個人に限定し、そのデータを予め蓄積し、これに合わせてパラメータを調整し、その特定される個人の覚醒度等心身状態を評価する発話音声分析装置を実現することは、現状の技術の延長上において十分に可能であると考えている。

しかしながら、筆者が実現を望む発話音声分析装置は、より平均的な人間像に対して適用可能な心身状態評価装置である。発話音声分析装置は、平均的な人間像を想定した旧来の予防安全装置へのアンチテーゼを目指した物ではあるが、完全に個人情報に依存する装置では世に役立つ物では有り得ない。発話音声分析装置における平均的な人間像は、時間や環境に対して常に変化する、その動物的・人間的なダイナミズムを平均的に有することを指しての平均的な人間像である。想定すべき平均的な人間像は、その人間の置かれた環境等、その状況から切り離されたものでは有り得ない。

発話音声分析装置の信頼性の向上においては、未だ明らかにしなければならない点が、ハードウェア、ソフトウェア、またその利用者として想定する人間の特性等において、無数に残されていると思われる。

これからも一つ一つの成果が積み上がる様に研究開発を続けて行きたいと考えている。

文 献

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 塩見, 他: 二種類の作業中発話音声のカオス性の変化, 日本人間工学会第 49 回大会, Jun. 2008.
- [5] <http://www.siceca.org>