カオス論的音声評価技術

○塩見 格一（電子航法研究所）

Chaotic Method for Voice Analysis

Kakuichi SHIOMI (Electronic Navigation Research Institute)

1. はじめに

従来の音声のスペクトラム分析が強度を色表示されたソノグラム上の3次元的特徴量（「語尾が上がっている」とか、「アクセントが前にある」とか、「音韻の変化に対してピッチが階段状に変化している」とか、等々のソノグラムを目で見れば理解できる特徴量、図1を参照）の評価であるのに対して、カオス論的な手法においては、より高次元空間に再構成された数学的なアトラクタにおいて実験的・経験的に定義される特徴量（「発話に係る筋肉の制御を埋込点の軌跡として表現している」と考えられるが、具体的に述べられることはない、多分、誰にも意味や内容は良くは分かってはいない特徴量、図2を参照）を評価する。

カオス論的な特徴量は高い自由度を有するので、評価目的に対して「適正な調整」が可能である場合には、ソノグラム上の特徴量の様な複雑な解釈を要せずに、評価対象尺度と単純に強く相関する特徴量を定義することも可能と思われるが、任意の評価尺度に対して上記「適正な調整」を実現する手法が明らかになっている訳ではない。

当所では、発話者の覚醒度（より正確には大脳の言語野の活性度）の評価を目的として研究開発を進めて来た。現時点において、覚醒度の定量化に十分有効と考えられる技術水準に達したので、これまでの経緯を含め、技術的な現状を説明する。

2. 研究開発の経緯と現状

従来、音声による喜怒哀楽の分析や、疲労状態の評価等は、多数の音声サンプルに対して評価目的に対するタグや指数値を与えることから始まり、人間の主観を評価指標とする他ないものであって、

客観性の確保が常に最大の問題であった。

上記のような従来手法に対して、当所手法においては、人間の耳で聞き分けられない特徴量を評価するものであるから必然的に、仮説検証型の実験を行うことにより、特徴量の適正化を目指してきた。

音声信号から算出するカオス論的な特徴量としては、リアプノフ指数の考え方を基本とし、これに時間局所性を加味した脳活性度指数（CEm/M: Cerebral Exponent micro/Macro）を定義し、これを算出するために SiCECA（Shiomi's Cerebral Exponent Calculation Algorithm）を開発した⁽²⁾。筆者は、「発話音声のカオス論的な特徴量を評価することにより、発話者の大脳新皮質の活性度が定量化できる」事を仮説として、様々な状況における音声を収録し、これら音声データから CEM 値を計算し、その性質を調べてきた⁽¹⁾。

鉄道総研との共同研究として実施した、被験者を極度な疲労状態に至らしめた鉄道車両運転操作模擬実験の結果において CEM 値の変化は、覚醒度の指標値としての臨界フリッカ周波数（CFF）の変化と高い相関を示した^(2,3)。

上記のような実験結果において、CEM 値には性差があることや、朗読テキストの音韻構成に依存すると思われる差異が存在すること等が明らかになり、また自動車や鉄道車両の運転等の主作業に伴って副作業として発話される音声からは、主作業負荷が著しく高い場合には CEM 値が低下することが観測され⁽⁴⁾、人間の脳の処理容量が有限である事に関わると考えられた。なお、航空管制業務のよりに主作業が知識集約型の作業であっても思考内容が高度に複雑化した様な場合には業務発話音声から算出される CEM 値は低下する⁽⁵⁾。覚醒度

図1 ソノグラム

各サンプル点の信号強度を右位相空間における埋込点の x 座標値として、これに対して τ 時間遅れたサンプル点の信号強度を同埋込点の y 座標値とすれば、埋込点の軌跡がストレンジ・アトラクタを描く。

図2 サンプリングされた時系列信号としての音声信号とターケンス・プロット

が高い状況における CEM 値の低下は、主作業と副作業に対する資源配分に依存すると考えられるが、これは業務姿勢に依るものであり、性格における個人差が認められる。

上記は実験結果において確認された人間の心身状態に係る CEM 値の性質であるが、CEM 値には不思議な統計的な性質があることも近年明らかになっている^(6, 7, 8)。

周波数分析であれば、同じ音声を同時に複数のレコーダで録音した場合、どの収録音声であっても殆ど変わらない結果が得られるが、SiCECA アルゴリズムにより処理すれば、同じ型式のマイクロフォンと同じ型式のレコーダを使用した場合であっても、大幅に異なる CEM 値が算出されることがある。この現象は、再生音声を再処理する過程で発見したものであり、5 秒程度の収録音声を 8 本のマイクロフォンを配置したスタンド前方 1～2m に設置したスピーカから繰返し再生し、収録した音声から算出される CEM 値の分布を調べた。

図 3 は、上記 8 本の内の 2 本のマイクロフォンにより収録した再生音声から算出した CEM 値の変化であり、M30-06 マイクロフォンにより収録した音声においては CEM 値が 3 図右のように双峰分布している。約 300 回の再生音声の収録は 5 時間程度の時間をかけて実施したものであり、その間の気温や湿度の変化による音速の変化等により、人間の耳では気付くことのない微妙な変化が双峰分布をもたらしたと考えられる。このような現象は、この実験まで気付かなかっただけであって、日常的に一定の確率で起きていた、と考えることが適当と思われる。血圧であれば、同じ安静状態で 5 回測っても、その内の 1 回の値が平均値に対して標準偏差の 3 倍も変化するようなことは通常はあり得ないと考えられるが、CEM 値については、2012 年当時の SiCECA の実装では、10 回計測すれば 1～2 回、平均値に対して標準偏差の 3 倍以上異なる値が日常的・平均的に算出される^(7, 8)。

3. おわりに

先の異常値の発生は CEM 値の信頼性を毀損するものであるが、複数のマイクロフォンで録音したり、複数の処理パラメータで並行処理したり、またオーバー・サンプリングによりデータを補間したりすることで、これを検出することは可能であり、2014 年に実現した SiCECA の実装においては、異常値の算出確率は 0.1% 以下に低減されている。

発話による発話者の覚醒度評価に、信頼性を向上させた SiCECA の利用を試みていただきたい。

参考文献

- 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. 4-1, pp3-12, Feb. 2004.
- 2) K. SHIOMI and et.al.: “Experimental Results of Measuring Human Fatigue by Utilizing Uttered Voice Processing” Proc. of IEEE SMC-2008 P0557, Oct. 2008, Singapore.
- 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：“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. 2006.
- 5) K. SHIOMI and et.al.: “Development and Evaluation of Fatigue and Drowsiness Predictor” Proc. of ICAS 2012, Sep. 2012, Brisbane.
- 6) 塩見：“発話音声のゆらぎの定量化における誤差について”，2009年電子情報通信学会総合大会，Mar. 2009.
- 7) 塩見：発話音声のゆらぎの定量化における誤差について-2，2010年電子情報通信学会総合大会，Mar. 2010.
- 8) 塩見：音声信号のゆらぎの定量化におけるマイクロフォンの特性，2011年電子情報通信学会総合大会，Mar. 2011.

図3 再生音声から算出された CEM 値の分布