

# 音声認識と画像認識を用いた疲労度推定システムの開発

## Development of a fatigue estimation system using voice and image recognition

西山 開貴  
Kaiki Nishiyama

津山工業高等専門学校 川波研究室  
Kawanami Laboratory, NIT, Tsuyama College

**概要** 日常を送るうえで疲労は避けられないものである。また、昨今の生活様式に伴い、電子機器の利用が増加している。これらの使用時間増加は、日常生活におけるさらなる疲労感の原因となる。そこで、疲労度を推定するシステムが有用ではないかと考え、疲労によって特徴が表れやすい表情と音声発話に着目した。本研究では、音声認識と画像認識の技術を用いて、ユーザがどの程度疲労しているかの推定を行い、必要に応じて休憩を促すシステムの開発を目指す。本稿では、システムの概要とデータ収録や、最終的な結果表示や休憩の促し方において検討したことについて報告する。

## 1 はじめに

人間は日常生活を送るうちに疲労する。疲労は疲労回復や疲労防止を心がけても避けるのは容易ではなく、疲労の蓄積のしやすさ、健康状態、職業柄などによる個人差はあるが、当人が気づいていなくても疲労していることは多い。また、昨今の情勢や情報化社会の影響でテレワークの普及、オンライン会議や授業の増加、ゲームのプレイ時間の増加などが挙げられ、それらを含めた電子機器の利用時間が増加傾向にある[1]。長時間画面を注視することで無意識のうちに疲労が蓄積しやすく、これによって眼精疲労や肩、首、腰の痛み、抑うつ症状をはじめとしたVDT症候群（テクノストレス眼症）や疲労蓄積のリスクが高くなり、日常生活にも影響が出ることが懸念される[2]。そのため、ユーザの疲労度を推定し、ユーザに通知するシステムが有用であると考えられる。音声発話や表情はユーザの状態によって特徴が表れやすく、日常的なコミュニケーションに用いることから特殊な訓練が不要である。また、それらを収録するためのマイクロフォンやWebカメラは安価に入手可能であり、ユーザは特殊な器具を着用する必要がない。また、安静にする必要もないため作業の中断がない[3]。特に音声はユーザの姿勢や照明の制限も受けない。

以上の点に着目し、図1に示すように機械学習を用いて

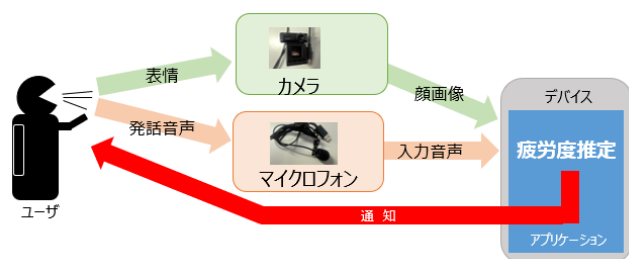


図1 システムの概要図

ユーザの発話音声と表情から現在の疲労度を推定し、疲労の度合いがあらかじめ定めたしきい値を超えた際、視覚的に分かるよう画面上に表示して休憩を促すシステムの開発を行う。先行研究を基にしながら、特に音声においては疲労等状態や生理状態の検出[4][5]、リアルタイム音声処理[6]、画像においては疲労検知や[7][8]、発話したユーザの顔の座標点を認識した特徴量抽出などの技術を応用して、システムの開発を目指す[9]。

最終段階においてはアプリケーションとして実装することを目指すため、先行研究の音声から疲労度を推定するアプリケーションの研究も参考にしながら[10]、入出力インタフェースの検討やシステムの評価、日常生活における実用化についても検討していく。

## 2 疲労について

### 2.1 疲労の定義

疲労とは、「過度の肉体及び精神活動、また疾病によって生じた独特の不快感や休養の願望を伴う身体の活動能力の減衰状態」であると日本疲労学会によって定義されている[11]。また、疲労には肉体的疲労と精神的疲労が存在するが、データ収録や疲労状態を再現する際の負担および研究背景を鑑み、本研究では後者を疲労状態として、音声、画像共に同条件で、状態の識別や疲労度の推定を行う。

### 2.2 通常状態

本研究における通常状態は、特に疲労を感じておらず、生体反応にも影響を及ぼしていない状態、また前述の疲労の定義に当てはまらない状態を示す。

### 2.3 疲労状態と状態の再現

本研究における「疲労状態」となる精神的疲労であるが、主となる原因は、社会生活や人間関係などのストレスによって生体機能への影響や、長時間のパソコン作業をはじめとする電子機器の長時間使用である。これらは視神経や脳の緊張状態から発生しやすい。また、精神的なストレスは即座に中枢系に働き、血圧の変動や心拍変動などを引き起こすこともある[12]。実際に、長時間パソコンに向かって作業している際に感じる疲れ、だるさは精神的疲労に分類される。

以上を踏まえ、本研究における疲労状態の再現は、PCでの研究開発作業を90分程度行い、100マス計算用紙を3枚解いた後の状態とする。

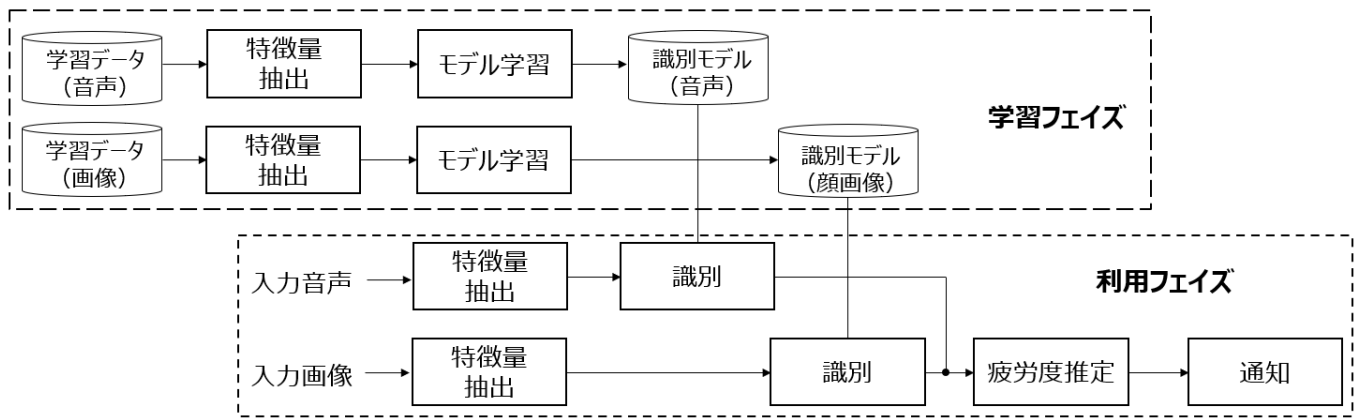


図2 処理の流れ

### 3 システムの概要

システムの処理の流れを図2に示す。使用者の発話音声や表情をマイクロフォンやカメラで読み込み、現在の疲労度をリアルタイムでユーザにわかりやすく画面等で通知することを目標とする。事前に、あらかじめ収録したユーザの音声と画像のデータを用いて、通常状態と疲労状態の識別モデルを作成する。なお、音声と画像では標準的な音声の分析フレーム間隔と画像撮影の間隔が著しく異なるため、それぞれ独立に特徴量を抽出して識別を行い、それらの識別結果から総合的な疲労度推定を行う。推定した疲労度に応じて、休息を促すなどのメッセージを付与してユーザに提示する。

ただし、モデル学習用データ、特に疲労状態のデータをユーザが事前に準備することはシステム利用の負担が大きい。そのため、話者非依存のモデルをユーザの通常状態と疲労状態のデータで逐次更新していくことが実用的システムでは求められるが、本研究では学習データが準備できるものとして、特定話者を想定して開発する。

### 4 疲労度推定手法

#### 4.1 音声の特徴量および疲労度推定手法

本研究はユーザの疲労度を推定することが目的である。状態の段階的識別を考慮すれば任意の多値分類器を用いることが可能であるが、ここでは筆者らの先行研究と同様[13]、音声や音の識別の基本的な手法となるGMM (Gaussian Mixture Models) の尤度や状態識別の結果を用い、画像における疲労度を推定するよう決定した。

また、特徴量も音声認識処理で標準的に用いられるパラメータ(表1)とした。これらは聴覚特性が考慮されているため身体の状態識別において最善であるかは不明だが、疲労の有無を人間でも声から知覚できる場合があることから、これを用いた。

表1 音声の特徴量

特徴量	MFCC(12), ΔMFCC(12), ΔPOWER(1)
混合数	64

#### 4.2 画像の特徴量および疲労度推定手法

画像において、使用する画像はユーザの通常状態または疲労状態の顔画像とした。Webカメラ (ELECOM社製 UCA M-C750FBBK) で撮影した画像から顔の座標点を抽出し、それらの特徴量としてニューラルネットワークによる学習を行った。最終的には識別結果と正解率から画像における疲労度を推定するように決定した。

### 5 データ収録とモデル学習

#### 5.1 データ収録

10代~20代の男性話者7名、女性話者1名で音声と画像の収録を行った。発話者からマイクは概ね300[mm]、カメラまでは概ね500[mm]となることを条件とした。収録環境を表2、収録装置概観を図3に示す。収録においては、通常状態と疲労状態で日を分けて行った。

音声の学習データはATR音素バランス文Aセット[14]、評価データは同Bセット冒頭の10文と日本語5母音の単独発声を収録した。なお、音素バランス文は各1回、5母音の単独発話は各2回とした。また、画像に関しては音声発話時に収録し、プログラム上で学習データと評価データを分けた。

表2 収録環境

話者	10代~20代男性7名、女性1名
収録環境	津山高専 卒研学生室
騒音レベル	24.1 [dBA]
収録ソフトウェア	WaveSurfer 1.8.8p4
標本化周波数	16000 [Hz]
量子化ビット	16
撮影環境	Webカメラから直接撮影
撮影間隔	30 [s]

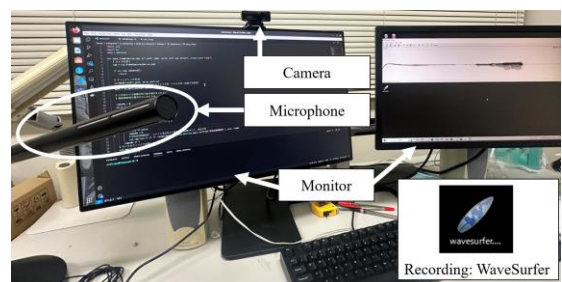


図2 収録装置概観

## 5.2 モデル学習および予備識別結果、考察

5.1節で説明した通り、現在は8名分のデータがあるが、まずは学習の前段階として、20代男声話者1名でモデル作成と予備識別を行った。

使用ツールとして、音声におけるモデル作成には音声処理ツールキットであるHTK (Hidden markov model ToolKit) を[15]、識別実験には汎用大語彙連続音声認識エンジンのJuliusを用いた[16]。画像では、読み込みや顔の認識にOpenCV[17]、モデル作成や機械学習にはscikit-learnを用いた[18]。画像に関しては自作プログラムを作成した。

実際に音声のモデルをHTKで作成し、Juliusにテストデータを入力して識別したところ、通常/疲労の識別はどちらも90%以上できており、正しく識別された際の尤度は少なくとも80%以上だったため、概ね識別できているのではないかといえる。疲労状態の条件は異なるが、同様の手法で以前筆者らが行った通常状態と肉体的疲労状態の予備識別実験よりも判別の精度、結果が良くなった[19]。

一方、画像においては学習データ12枚、テストデータ4枚の計16枚を用意し、複数回学習を実行した結果の正解率は50~100%となった。現段階では予備識別のためデータ数が少なく、正解率が安定していない結果となった。今後に向けて、音声においては発話困難者や不特定多数の使用を考慮し、学習データやテストデータの見直し、話者非依存のモデルに関する検討、現在のモデルを用いながらも話者に依存せず動作させることについて検討が必要である。

画像においては、正解率が100%となるケースも見られたことや、既にもとの画像は収録されているため、処理しデータ数を増やすことで、さらに安定した結果が得られるのではないかと考えられる。また、6節において説明するが、音声と同様に、通常状態か疲労状態の2値識別も実装する必要がある。

## 6 疲労度推定結果に対する検討

本研究においては、疲労の度合いをどう決定していくかが今後の重要な要素である。本研究ではユーザの発話音声と表情から現在の疲労度を推定し、疲労の度合いがあらかじめ定めたしきい値を超えた際、視覚的に分かるよう画面上に表示して休憩を促すシステムの開発を行うことが目的である。

よって、最終的な疲労度推定結果について、識別結果(通常または疲労の2値分類)および音声の尤度と画像の正解率を足して2で除した結果を「総合推定結果」として以下(1)~(6)に分類する(なお、尤度と正解率は同じものとして計算する)。結果が(1)~(3)の場合は休憩を促す通知を出す。以上を一検討案として挙げる。

なお、この場合は音声か画像のどちらかの識別結果が著しく偏る場合や、片方は通常だがもう片方は疲労状態となる場合が考えられるため、その場合は(4)を結果として表示する。この場合も休憩を促す通知を出す。

### 疲労度推定総合結果一覧

- (1) 完全な疲労状態である
- (2) 疲労の可能性が高い
- (3) 多少の疲労が見られる
- (4) 疲労の可能性はある
- (5) 疲労の可能性は低く、ほぼ通常状態である
- (6) 完全な通常状態(元気)である

## 7 まとめ

本稿では、疲労度を推定するシステムの開発を行うことを目的として、提案手法を基にデータ収録やモデル作成、予備識別実験、評価、考察を行った。今後は収録したデータを基にしながらモデルを作成し、総合的に疲労度を推定できるようにする。また可能であればデータ数を増やし、より信頼性のあるシステムに向けて開発を進めていく。

## 8 参考文献

- [1] 森健, 林裕之, “新型コロナウイルス感染拡大で生活におけるデジタル活用が急進展,” 野村総合研究所, p p.1-2, May. 2020.
- [2] 延知奈美他, “VDT機器を使用した学習環境における疲労検知のための自己接触動作認識,” 教育システム情報学会誌, Vol.35, no.2, pp.145-148, Apr. 2018.
- [3] 松村寿枝他 “音声分析を用いた自転車シミュレータ運転時の疲労測定法の基礎的検討,” 電気学会論文誌C (電子・情報・システム専門誌), vol.136, no.1, pp.94, Jan. 2016.
- [4] 有賀康顕他, “生理情報を基にした音声による緊張程度検出に関する検討,” 23rd Fuzzy System Symposium, pp.597-601, Aug. 2007.
- [5] 北越大輔他, “音情報を利用した感情・疲労度推定システムの構築: 感性情報処理システムへの拡張に向けて,” サテライト・ベンチャー・ビジネス・ラボラトリ一年報, Vol.5, pp.52-54, 2003.
- [6] 塩見格一他, “疲労等判定用発話音声分析装置の試作開発,” 人間工学 第43巻特別号, pp.198-199, Jun. 2007.
- [7] 黒澤勇樹他, “機械学習を用いた口唇動作による精神疲労推定手法,” パーソナルコンピュータ利用技術学会論文誌, vol.14, no.1, pp.22-28, Mar. 2020.
- [8] 野々村拓朗他, “Webカメラ画像の多次元解析による疲労推定システムの研究開発,” 情報・システムソサエティ特別企画 ジュニア&学生ポスターセッション予稿集, pp.166, Mar. 2021.
- [9] 駒井祐人他, “AAMを用いた唇領域特徴による音声発話認識,” 電子情報通信学会技術研究報告=IEICE technical report: 信学技報109 (374), pp.357-362, Jan. 2010.
- [10] 青木由希他, “音声から疲労程度を推定するスマートフォン用アプリケーションの開発,” FIT2013 (第12回情報科学技術フォーラム), 第2分冊, pp.269-270, Sep t. 2013.

- [11] “抗疲労臨床評価ガイドライン,” pp.4, (社) 日本疲労学会, Dec. 2011.
- [12] 南谷晴之, “疲労とストレス,” バイオメカニズム学会誌, Vol.21, no.2, pp.58-59, May 1997.
- [13] KEODUANGDY NITIPHON, 川波弘道, “音声を用いた疲労度推定に関する研究,” 電子情報通信学会情報・システム講演論文集1, pp.94, Mar. 2023.
- [14] 匂坂芳典他, “ATR音声・言語データベース,” 日本音響学会誌, 48巻, 12号, pp.878-882, Dec. 1992.
- [15] “HTK Speech Recognition Toolkit”, <https://htk.eng.cam.ac.uk/>.
- [16] “大規模連続音声認識エンジン Julius”, <https://julius.osdn.jp/>.
- [17] “OpenCV – Open Computer Vision Library”, <https://opencv.org/>.
- [18] “scikit-learn : machine learning in Python – scikit-learn 1.3.0 documentation”, <https://scikit-learn.org/stable/>.
- [19] 西山開貴, 川波弘道, “音声を用いた疲労度推定に関する研究,” 音声コミュニケーション研究会資料, vol. 3, no.4, pp.10-11, Sept. 2023.