

# 演技発話による疲労の表現によって生じる音色変化の分析

生野 琢郎<sup>†</sup> 森勢 将雅<sup>††</sup>

<sup>†</sup> 山梨大学 大学院 医工農学総合教育部 〒400-8511 山梨県甲府市武田 4-3-11

<sup>††</sup> 山梨大学 大学院総合研究部 〒400-8511 山梨県甲府市武田 4-3-11

E-mail: <sup>†</sup>{g17tk010,mmorise}@yamanashi.ac.jp

あらまし 音声信号の疲労を示す生体指標としての可用性を調査するために、通常の発話と疲労状態を演技した発話をそれぞれ収録し解析した。日本語の母音5種類をそれぞれ含む5文章を用いて、男女各5名による発話を収録した。解析は、音色に関係する音響特徴量である、スペクトル重心、スペクトル傾斜、非周期性指標、第1フォルマント、第2フォルマントに対して行った。各特徴量について、通常発話の時の特徴量に対する疲労演技の際の特徴量の変化率を算出し、通常時と疲労演技時の音響特徴量を比較した。比較の結果、スペクトル重心は演技によって低下し、非周期性指標は増加する傾向が確認された。

キーワード 音声知覚, 音声分析, 疲労, 音響特徴量

## Analysis of timbre changes caused by expressing fatigue speech

Takuro SHONO<sup>†</sup> and Masanori MORISE<sup>††</sup>

<sup>†</sup> Integrated graduate school of medicine, engineering, and agricultural sciences, University of Yamanashi, 4-3-11, Takeda, Kofu-shi, Yamanashi, 400-8511 Japan

<sup>††</sup> Integrated Graduate School, University of Yamanashi, 4-3-11, Takeda, Kofu-shi, Yamanashi, 400-8511 Japan

E-mail: <sup>†</sup>{g17tk010,mmorise}@yamanashi.ac.jp

**Abstract** To examine whether speech can be used as a biomarker for human fatigue, we recorded both usual speech and speech by acting fatigue and analyzed them. Five men and women uttered five short sentences including five Japanese vowels respectively. Recorded speech was analyzed by using acoustic features related to timbre (spectral centroid, spectral tilt, aperiodicity at 3 kHz, the first and the second formant frequencies). We compared these features in usual speech with that features in acted speech by calculating change ratio. As a result of comparison, spectral centroid tends to decrease, and the aperiodicity tends to increase by acting fatigue.

**Key words** speech perception, speech analysis, fatigue, acoustic feature.

### 1. はじめに

労働環境などから来る慢性的な疲労やストレスが蓄積し、心身に不調をきたすケースがある。平成27年の段階で、メンタルヘルスケアに取り組んでいる事業所の割合は、増加している。仕事や職業生活に関する強い不安、悩み、ストレスを感じる労働者の割合は半数を超えている[1]。このようなことから、現在、疲労やストレスのケアに対する関心が高まっている。現在、自身のストレスや疲労を定期的に評価する方法として、自覚症しらべのような質問紙への回答が挙げられる[2]。一方、質問紙への回答は、回答への時間を要する為、定期的な検診には時間がかかることが予想される。そこで本研究では、非接触、かつ簡易的に計測が可能な音声信号における、人間の疲労度に対応す

る特徴量の調査を目的とする。前報では、被験者3人による疲労を演技した音声の解析を行い、傾向を確認した[3]。本稿では、男女5名の発話者による演技音声を収録・解析し、統計的な解析を行う。感情音声では演技発話と自然発話との間に音響特徴量の差が生じることが指摘されている[4]。本来であれば実疲労を伴う音声を解析すべきであるが、疲労の程度や種類にも多様性があることを勘案し、初期的な検討として演技疲労を用いた解析を行う。この解析結果から疲労表現に関連する音響特徴量を策定し、実疲労を計測するための研究に役立てる指針で研究を進める。

本稿では、第2章において本研究の位置づけを関連研究と共に示す。第3章では、本稿で行った演技疲労発話の収録実験、収録音声から音響特徴量を抽出する方法について説明する。第

4章では、通常音声と疲労演技音声の比較を行う手法とその結果を示す。第5章で結論を述べる。

## 2. 関連研究と本研究の位置づけ

疲労度を定期的に検診するための手法として、人間計測によって得られる生体指標を用いる手法が挙げられる。主な研究事例として、唾液を採取し、そのコルチゾール濃度を調べることで急性ストレスに対する反応の計測が可能であり、慢性ストレスについてはヘルペスウイルスの再活性化反応が疲労に関係することが示唆されている[5],[6]。一方、これらの手法は唾液を採取し調査を行う必要があるため、計測の際に人体への接触を要し、検査キットの定期的な購入にコストがかかる。

近年の音声処理技術の発展により、音声の発話内容以外の情報を認識する研究が多く報告されている。悲しみ、喜び、怒りといった基礎的な感情の認識の試み[7],[8]はもちろん、今回着目するストレスや疲労と言った複雑な感情や心理現象についての検討も行われている。松村らは、3種類の文章の発話に対して、聴覚実験より疲労と非疲労を分類し、それぞれの比較を行った。また、VDT (Visual Display Terminal) 作業を行う前後の音声を3種類の文章について収録し、解析を行った[9]。結果、疲労と判定された音声や、負荷を与えた後の音声については、平均パワーや継続長、基本周波数と言った基本的なパラメータが変化する傾向が示唆された。また、Krajewskiらは、夜間に睡眠をとらないで動画を見た際の、眠気に関連する音響特徴量の調査、識別器による眠気の2値判定評価を行った。結果、眠気を示す指標とHNR (Harmonics-to-noise ratio)を含む5つの音響特徴量は他の特徴量に比べて高い相関がみられるという結果が得られた[10]。

その一方、疲労の表現に関して検討した事例は少ない。酒造らは、感情認識に用いる為の、様々な感情を有する自然発話と演技発話を多数含む感情音声データベースの作成を行った。発話音声を聴取し、発話の感情や情動を回答する主観評価実験を行った結果、演技発話音声に対する感情の誤認識率は5%を下回っていた[11]。疲労感によって人間の音声に変動が生じると仮定すると、疲労感を演出した音声を発話することも可能であることが考えられる。ここで、音声信号を生体指標として用いるのであれば、主観的な疲労と身体の状態としての疲労を明確に区別できなければならない。疲労に関して意図的に操作可能な音響特徴量がどのようなものかを明らかにしておくことは、身体の状態としての疲労を解析する際に有効な知見であると考えられる。

そこで本稿では、日本語5母音をそれぞれ含む5つの相槌文を、通常の発話と疲労を演技して発話した場合での音響特徴量の比較を行った。現在疲労が人間に与える影響という点では明らかになっていない点が多い。本解析によって、疲労に対して人間が主観的に操作可能なパラメータを明らかにすることを目的とする。

## 3. 疲労を演技した音声の収録と音響特徴量抽出

疲労演技による音響特徴量への影響を調査するために、男女

表1 音声収録の環境と収録文章

Table 1 Recording condition and sentences used in the evaluation

発話環境	防音室
騒音レベル	18 dB (A-weighted SPL)
録音機材	EG2-PCX2, M80 (Glottal Enterprises, Inc.)
サンプリング	44.1 kHz / 32 bit
発話文章	「あぁ」「いいね」「うーん」 「えーっと」「おお」の計5種類
発話者	20代前半の男性3名

各5名の発話者による5種類の相槌文の発話を、通常時と疲労を演技した時の2通りについてそれぞれ収録し、音声解析を行った。本章では、音声の収録内容と収録方法、収録した音声から音響特徴量を抽出する方法について説明する。

### 3.1 発話する文章とその収録条件

本研究で用いた音声の収録条件を表1に示す。男女各5名からなる大学生、大学院生を発話者とした。発話に用いた文章は、日本語5母音をそれぞれ含む、日常的に良く発話されると考えられる相槌文5種類である。発話者は各文章について、通常の状態での発話を5回、疲労状態を演技した状態での発話を5回、計10回の発話を行う。発話者には、各文を発話する際の仮想的な状況についてそれぞれ説明し、説明した状況下を想定して発話をするように依頼した。一例として、“あー”という発話に対しては、相手の言ったことに同意を示すようなイントネーションで読むように指示した。今回用いた各文章は発話時間が3秒以内の長さである。また、発話時の状況を指定していることから、音量によるイントネーションについて、発話者間の個人差をある程度抑制できると考えられる。また、各文章は5つの母音をそれぞれ文頭に含むため、疲労演技による音響特徴量への影響における、母音に対する依存性の分析を可能にする。

### 3.2 音響特徴量の抽出方法

収録音声の解析は次の手順で行った。まず、収録された5種類の発話を、音素系列がそれぞれ、/a/, /i/, /u/, /e/, /o/, となるように音声抽出した。抽出した音声を、音声分析合成システム WORLD [12] と音声解析ソフト Praat を用いて解析した。各音声信号を16 kHz にダウンサンプリングし、音声信号の時間フレームごとのスペクトル包絡、3 kHz の周波数成分における非周期性指標を D4C [13] によって算出する。非周期性指標は、声のかすれ具合を表すパラメータである。また、スペクトル包絡の1 kHz から8 kHz までに最小二乗法を適用する事でスペクトル傾斜を算出した。スペクトル傾斜は、音の鋭さを示すパラメータである[14]。そして、スペクトル包絡の0 Hz から8 kHz までの成分に、式(1)を用いてスペクトル重心[15]を算出する。

$$C_n = \frac{\sum_{k=0}^N \omega_k |X_n(\omega_k)|}{\sum_{k=0}^N |X_n(\omega_k)|}, \quad (1)$$

ここで、 $C_n$  は、音声の  $n$  番目の時間フレームにおける、スペクトル重心を表す。 $N$  は、8 kHz を指し示す離散周波数フレームを示す。 $|X_n(\omega)|$  は  $n$  番目の時間フレームにおける、周波数  $\omega$  のパワーを示している。また、Praat を用いて、第1フォルマント周波数 (F1) と第2フォルマント周波数 (F2) をそれぞれ

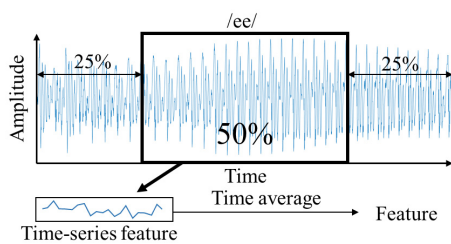


図1 音響特徴量の算出の手順

Fig. 1 Procedure in calculating acoustic features

算出した。フォルマントは、声道の形状によって定まるスペクトル包絡上のピークである。F1, F2 はそれぞれ顎の開き, 舌の形状によって大きく変化する事が示唆されている [16].

図1に、1回の発話に対する各特徴量の抽出方法を示す。まず、各発話に含まれる文頭の母音の開始・終了にラベルを付与する。ラベリングされた区間の内、中央の50%の区間に対して、前述の各特徴量をフレーム毎に算出し、時間方向の平均を算出する。この処理によって、1回の発話に対して各特徴量が1つずつ得られる。得られた各特徴量を発話回数5回について平均することで、通常の発話、疲労を演技した発話の各母音について各特徴量が得られる。

#### 4. 解析結果・考察

本章では、抽出した音響特徴量の統計解析の方法を説明し、通常の発話に対して、疲労を演技した発話の音響特徴量がどれ程変化するかを述べる。そして、結果から考察を示す。

##### 4.1 解析結果の統計解析

通常の発話と疲労を演技した発話における各特徴量の比較方法について説明する。まず、実数で扱う F1, F2, スペクトル重心, スペクトル傾斜の比較には、式 (2) を用いた。ここで、

$$C_r = E(X_{act})/E(X_{plain}), \quad (2)$$

$E(X_{plain})$ ,  $E(X_{act})$  は、前述の処理で得られた、疲労を演技した発話時と通常の発話時の特徴量をそれぞれ示している。そして、対数で扱う 3 kHz 帯の非周期性指標に関しては、次の式 (3) を用いた。

$$C_r = E(X_{act}) - E(X_{plain}), \quad (3)$$

式 (3) によって、対数においても実数と同様の変化率を算出する。この各式によって、通常時の発話に対して、疲労を演技することによって生じる音響特徴量の変化の割合を算出する事が出来る。特に、フォルマント周波数に関しては声道長による男女差が存在するが、変化率を見ることで男女の絶対的なフォルマント周波数の差を無視できると考えられる。

##### 4.2 解析結果

図2, 3, 4に、今回検討を行った5つの特徴量のうち、F1, スペクトル重心, 3 kHz の非周期性指標の解析結果を示す。グラフ中の横軸は、該当する特徴量における、疲労演技による通常の発話からの変化率に対する平均とその95%信頼区間を示す。実数で扱った F1, スペクトル重心に関しては、1であれば変化は無く、1より大きければ疲労演技によって増加、1より小さければ疲労演技によって低下するという傾向を示している。対

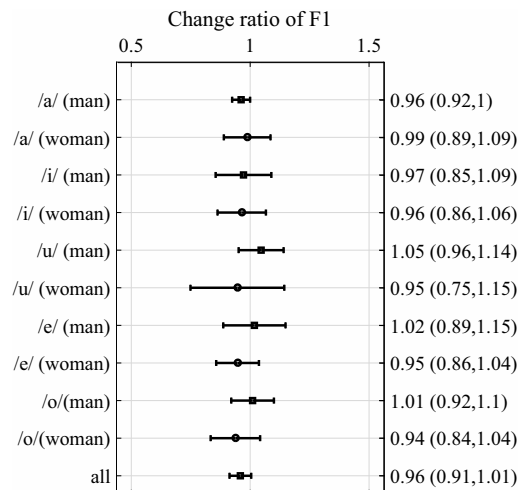


図2 母音、性別毎の変化率。各プロットは平均とその95%信頼区間を示す。有意な傾向は見られなかった。

Fig. 2 Change ratio of the F1 in each vowel and gender. Each plot shows average and 95% confidence interval. We couldn't confirm significant tendency.

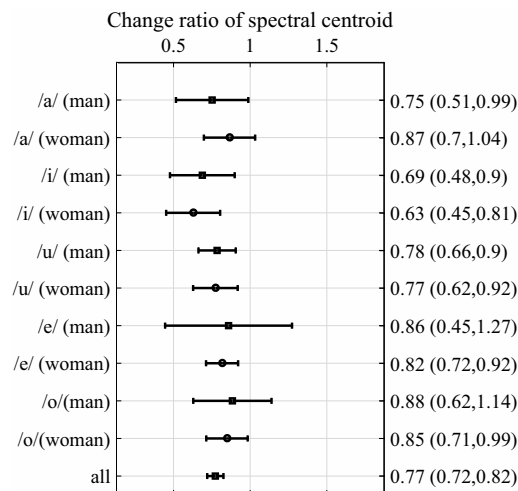


図3 母音、性別毎のスペクトル重心の変化率。全体としてスペクトル重心は減少傾向にあるという事が確認された。

Fig. 3 Change ratio of the spectral centroid corresponding to each vowel and gender. We confirmed that the spectral centroid tended to decrease by acting.

数で扱った非周期性指標に関しては、0であれば変化が無く、0より大きければ疲労演技によって増加するという傾向を示している。また、縦軸左側は母音と性別に関するラベルを示している。最下段にある all は、性別、母音を考慮しない全体の傾向である。そして、縦軸右側は変化率の平均値と、その95%信頼区間の範囲を数値として示している。

図2より、F1の変化率は母音、被験者にかかわらず全体として1以下であるが有意ではない事が確認された。また、図3より、母音/e/, /o/における男性発話と、母音/a/における女性発話を除き、スペクトル重心は疲労演技によって低下する傾向が確認された。図4に示すように、非周期性指標による疲労の表現は、発話者や文章による変動が大きく見られるが、全体としては増加傾向にある事が確認された。本稿では紙面の都合上割愛したが、スペクトル傾斜, F2については被験者による変化率の個人差が大きく、有意な傾向も見られなかった。

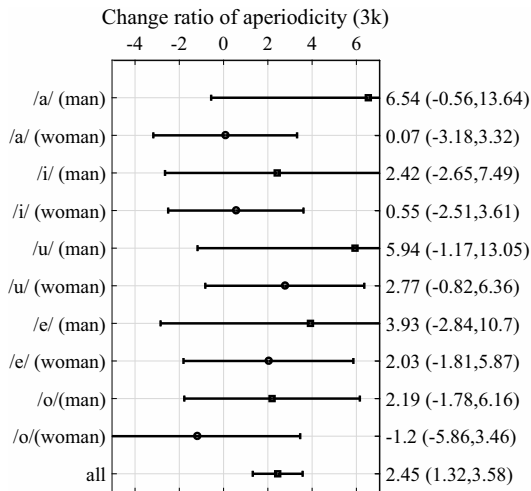


図4 母音、性別毎の3 kHzにおける非周期性指標の変化率。非周期性指標の変化はばらつきが大きいですが、全体の傾向としては増加傾向にある事が確認された。

Fig. 4 Change ratio of the aperiodicity in 3 kHz corresponding to each vowel and gender. The aperiodicity tends to increase by calculating the whole result, while the individual differences are large.

#### 4.3 考察

今回検討した5つの特徴量のうち、スペクトル重心は、声の明るさに対応するパラメータである[15]。したがって発話者は、疲労を意識した際に声の明るさに注意を向けて発話する事が示唆された。また、F1は変化率が1近傍に分布しているが、減少傾向にある事が推測される。先行研究において、体調の変化に関連する変化として、体温の低下に伴う声道壁の熱伝導の変化が声道の伝達関数を変化させ、フォルマント周波数を低下させることが示唆されている[10]。そのうち、F1は顎の開閉に関連すると言われている。従って、顎の開閉状態の変化はあったが、声道壁の状態などの細微な部分による音響特徴量の操作が出来なかった可能性が考えられる。非周期性指標に関しては、疲労音声の演技について、ため息が混ざったような発音をする発話者が見られたため、その影響が大きいと考えられる。

人間は、状況によって発話の音色を変更することが可能である。複数の発話者から共通の傾向が認められた事から、話者が疲労を表現した発話をする際、吐息交じりの、暗い音色を有する音声を意図する可能性が示唆された。今後は、収録音声に対する主観評価実験により、知覚される疲労度と演技表現の対応に関する検討を行う。

#### 5. おわりに

本研究では、男女各5名による通常時の発話と、疲労を演技した発話を分析し、疲労演技による音色に関する音響特徴量の変化を調査した。調査の結果、母音、被験者に依らずスペクトル重心は低下し、非周期性指標は増加するという傾向が見られた。この結果より、疲労の表現に際し発話者は、声の明るさやかすれ具合に対する意識を向けるという事が示唆される。

今後の展望としては、今回収録された音声に対し主観評価実験を行い、疲労を他者に伝える際の表現と、聴取者の知覚に関

する対応の検討が挙げられる。主観評価実験によって、今回行われた表現が、知覚に基づく表現であったのか、イメージによって行われたものであったのかを調査する事が必要である。また、今回は疲労に関連する操作可能な音響特徴量に関しての検討を行った。この知見を基に、疲労によって意図せずに変化する音響特徴量を調査し、生理学的・心理学的に考察することが課題となる。

#### 6. 謝辞

本研究は、科研費 JP15H02726, JP16H05899, JP16K12511, JP16H01734 の支援を受けて実施された。

#### 文献

- [1] 厚生労働省, “平成 29 年版過労死等防止対策白書,” pp.18–24, 2017.
- [2] 久保 智英, 城 憲秀, 武山 英磨, 榎原 毅, 井上 辰樹, 高西 敏正, 荒鷹 優子, 村崎 元五, “「自覚症しらべ」による連続夜勤時の疲労感の表出パターンの検討,” 産業衛生学雑誌, vol.50, no.5, pp.133–144, 2008.
- [3] 生野 琢郎, 森勢 将雅, “音声の疲労度に対応する音響特徴量に関する基礎的検討-演技された音声をを用いた分析-,” 日本音響学会 2017 年秋季研究発表会講演論文集, pp.335–336, 2017.
- [4] R. Jurgens, K. Hammerschmidt and J. Fischer, “Authentic and play-acted vocal emotion expressions reveal acoustic differences,” *Front Psychol.*, vol.2, no.180, pp.1–11, 2011.
- [5] 田中 善秀, 脇田 慎一, “ストレスと疲労のバイオマーカー,” *日本薬理学雑誌*, vol.137, no.4, pp.185–188, 2011.
- [6] 近藤 一博, “HNV-6 の潜伏感染・再活性化のバイオマーカーとしての有用性,” *日本補完代替医療学会誌*, vol.3, no.2, pp.61–67, 2006.
- [7] 直井 克也, 松本 哲也, 竹内 義則, 工藤 博章, 大西 昇, “感情に関係する特徴量の検討,” *電子情報通信学会技術研究報告, HIP, ヒューマン情報処理*, vol.105, no.99, pp.37–42 2005.
- [8] T. Polzehl, A. Schmitt, F. Metze and M. Wagner, “Anger recognition in speech using acoustic and linguistic cues,” *Speech Communication*, vol.53, no.9–10, pp.1198–1209, 2011.
- [9] 松村 寿枝, 吉村 宏紀, 清水 忠明, “音声分析を用いた話者疲労の推定,” 2010 年電子情報通信学会総合大会論文集, vol.143, p.143, 2010.
- [10] J. Krajewski, R. Wieland and A. Batliner, “An acoustic framework for detecting fatigue in speech based human-computer-interaction,” *Computers Helping People with Special needs, ICCHP*, pp.54–61, 2008.
- [11] 酒造 正樹, 山本 泰史, 志村 誠, 門間 史晃, 光吉 俊二, 山田 一郎, “情動・感情判別のための自然発話音声データベースの構築,” *情報処理学会論文誌*, vol.52, no.3, pp.1185–1194, 2011.
- [12] M.Morise, F. Yokomori and K. Ozawa, “WORLD: a vocoder-based high-quality speech synthesis system for real-time applications,” *IEICE Transactions on Information and Systems*, vol.E99-D, no.7, pp.1877–1884, 2016.
- [13] M. Morise, “D4C, a band-aperiodicity estimator for high-quality speech synthesis,” *Speech Communication*, vol.84, pp.57–65, 2016.
- [14] 岩宮 眞一郎, “図解入門よくわかる最新音響の基本と応用,” pp.94–97, 秀和システム, 2011
- [15] E. Schubert and J. Wolfe, “Does timbral brightness scale with frequency and spectral centroid?,” *Acta Acustica United with Acustica*, vol.92, pp.820–825, 2006.
- [16] J. Sundberg, “The science of the singing voice,” 1987. 榑原 健一, 伊藤 みか, 小西 知子, 林 良子: 歌声の科学, pp.93–134, 東京電機大学出版局, 2007.