

令和元年6月21日現在

機関番号：30110

研究種目：基盤研究(B) (一般)

研究期間：2015～2018

課題番号：15H03207

研究課題名(和文)生成メカニズムに基づく声質の音声学的分類

研究課題名(英文)Phonetic classification of voice qualities based on production mechanisms

研究代表者

榊原 健一 (SAKAKIBARA, Ken-Ichi)

北海道医療大学・リハビリテーション科学部・准教授

研究者番号：80396168

交付決定額(研究期間全体)：(直接経費) 12,600,000円

研究成果の概要(和文)：喉頭音源に関連する声質の客観的表記に関し、生理学的、物理学的視点から、音源となる振動体の分類により(i)声帯発声；(ii)声帯-仮声帯発声；(iii)声帯-披裂部発声；の3種類の有声音と、無声音である声門乱流雑音発声と分類が可能であり、有声音源に関しては、物理的・音響的特徴から周期的、準周期的、非周期的の3つの振動への分類を提案した。また、声質表現として緊張-弛緩と声門開放時間率との関係を明らかにし、声質の客観的な表現語として適切であることを明らかにした。  
声門開放時間率の分析方法として、余弦級数包絡の複素wavelet分析に基づく簡易なGCI、GOIの検出方法を考案した。

研究成果の学術的意義や社会的意義

声質に関しては主観的な定義の曖昧な表現語が用いられてきたが、本研究成果により、客観的に解釈、理解可能な声質表現が実現されたことは、声質関連研究において共通理解が可能な用語が確立されたという学術意義がある。また本研究を遂行する過程で、提案された様々な声質に関連した分析方法は、今後の音声分析・合成のために用いることが可能であり、感情分析、感情音声合成など具体的な応用が期待される。

研究成果の概要(英文)：As the objective classification of voice quality of voiced sounds related to the laryngeal source, we proposed four different categories concerning the vibration portion in the larynx: (i) vocal fold phonation; (ii) vocal-ventricular phonation; (iii) vocal-aryepiglottic phonation. Also, we described unvoiced sound source as glottal turbulence. For voiced sounds, it is possible to use physical and acoustic features, and periodic, subharmonic, aperiodic. Besides, we clarified the relationship between tension-relaxation and open quotient of vocal fold vibration as voice quality expression, and clarified that it was suitable as an objective expression word of voice quality.

We developed a simple GCI and GOI detection method based on complex wavelet analysis of cosine series envelope as an analysis method of glottal open time rate.

研究分野：音声学

キーワード：声質 音響分析 音声記号

## 様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

### 1. 研究開始当初の背景

声質は、音声 = 話し言葉における声の音色のうち、音韻性に関与しない諸要素である。話し言葉において、声質は、意図、感情、発話様式などパラ言語情報ならびに性別、年齢などの非言語情報を担い、人間の音声コミュニケーションにおいては、言語情報とともに重要かつ本質的な役割を果たしている。

音声の言語情報は、音声学的には、IPA (国際音声字母) により記述する。世界中の言語の音韻体系は、母音、子音とも、発声および調音運動、聞こえに基づいた調音音声学的な IPA による記述によって体系化されている。

しかしながら、声質は、言語における音素とは異なり、連続的な音響的特徴を持ち、音素のように明確な音響ターゲットが存在しない場合が多く、そもそも記号による記述が困難で、これまで声質を記述する様々な方法が提案されてきたが、それらの記述法の多くは、聴覚印象に基づく印象語による記述で曖昧で、普遍性に欠けるといふ記述体系としての決定的な問題点が存在する。

Laver は生成メカニズムに結び付けた語を用い、聴覚印象により声質を分類し、著書に音声資料とともに声質分類の体系を記述した[1]。しかしながら、これら声質の分類は、生理学的観察に基づくものではなく、声質の生理学的な生成メカニズムの違いに関する混乱を招き(例えば、harsh は生理学的には声帯-披裂喉頭蓋発声であり[2]、creaky-falsetto は生理学的には不可能など) また、普遍性に欠け記述体系として研究者間で共有されないまま現在に至っている。

[1] J. Laver, *The Phonetics description of voice quality*, Cambridge Univ. Press. 1980.

[2] K.-I. Sakakibara, *Production mechanism of voice qualities in singing*, *J. Phonetic. Soc. Jpn.*, 7(3):27—39, 2003.

### 2. 研究の目的

本研究は、声質の生成メカニズムを生理学的観察により明らかにし、生成メカニズムと音響的特徴とを関連付けることによって、定量的な分析に基づく客観的かつ普遍性をもつ声質の体系的な音声学的分類方法を構築することを目的とした。

従来の声質の音声学的分類は、声質の音響的連続性、音声生成の複雑性を考慮せず、主として聴覚的印象評価により、分類からは発声の特徴が同定できない曖昧な定義が用いられてきた。

本研究では、病的音声、感情音声、歌声を含む様々な音声の声質について、声帯振動や喉頭の動態、舌運動を、生理学的観察および音響分析により明らかにし、また感覚運動の視点から聴覚印象評価と生成メカニズムを結び付け、客観的かつ体系的な声質の音声学的分類方法を構築することを目標とした。

本研究では、声質の音声学的記述という課題を解決するために、音素の音声記述と同様、声質の生成メカニズムと、音響的特徴とを関連付けることにより、客観的かつ普遍性をもつ声質の音声学的な記述の体系、分類方法を構築することを目的とした。

### 3. 研究の方法

普遍性かつ客観性をもつ声質の記述には生理学的な視点が不可欠であり、音声生成の視点からは、主に喉頭調節により制御される喉頭音源(声帯、仮声帯、披裂喉頭蓋の振動および喉頭室共鳴など)、下咽頭(梨状陥凹、喉頭腔の形状による音響的影響)、言語に依存するが非言語情報に関与する声道形状の変化(鼻腔の音響カップリングなど)、更には、声道と音源の非線型相互作用により規定される。本研究では、声質を生成する発声運動を生理学的観察方法で観測し、また、声質に関する音響分析を用いて解明し、IPAの音素記述と同様に生理学的な用語による普遍性かつ客観性を持つ声質の音声学的記述体系を構築をおこなった。

### 4. 研究成果

#### (1) 喉頭音源に関連する声質の客観的表記

喉頭音源の生理学的、物理学的視点からの声質の分類基準は、まず、音源となる振動体の分類により(i) 声帯発声(vocal fold phonation); (ii) 声帯-仮声帯発声(vocal ventricular phonation); (iii) 声帯-披裂部発声(vocal-aryepiglottic phonation); の3種類の有声音と、無声音である声門乱流雑音発声(glottic turbulent noise phonation)と分類ができる。声帯発声、声帯-仮声帯発声、声帯-披裂部発声の有声音源に関しては、物理的・音響的特徴から周期的(periodic)、準周期的(subharmonic)、非周期的(aperiodic)の3つの振動に分類可能である。

また、従来から声質表現に用いられる緊張-弛緩(tense-lax)の概念は、生理学的な筋緊張-筋弛緩の喉頭調節に対応した概念であるが、物理的には声帯振動の声門開放時間率( $O_q$ )、また音響的には $O_q$ と強い相関のあるスペクトル傾斜によって客観的に評価することが可能であり、声質の客観的な表現語として適切である。

また、発声の基本周波数に関連した生理学的に異なるメカニズムによって実現される声区は、これまでの研究で、生理学的視点から甲状披裂筋支配の声区(TA-dominant register)と輪状甲状筋支配の声区(CT-dominant register)とに分けることが可能で、それぞれの声区に緊張-弛緩の尺度を組み合わせることで、従来の声区を、フライ声区(vocal fry)を low-pitched lax TA-dominant、ホイ

ッスル声区を high-pitched tense CT-dominant のように表現することが可能である。

声質の定義からすると、声の基本周波数は声質とは独立であるが、音声学表記における喉頭音源の表現には、声質の多くの要素が基本周波数によって制約、特徴づけされるので、基本周波数の要素を high-pitched, middle-pitched, low-pitched などの表現で記述することは必要である。

## (2) 声門開放時間率( $O_q$ )の分析方法の提案

声門開放時間率は、声帯振動に関連する特徴量で、周期内での声門が開放されている時間の比率を示す。声門開放時間率が、発声様式と関連することが、従来研究でも報告されていた。ただし、声門開放時間率の実際の計算には、様々な課題が存在した。

声門開放時間率は、(i) 電気式声門図(Electroglottography; EGG)の波形から計算する方法; (ii) 声帯振動の高速デジタル撮像から計算する方法; (iii) 逆フィルタによる音源堆積流から計算する方法; が有力な方法であるが、(i) では、閉鎖イベントと(GCI)とか開大イベント(GOI)の検出において、EGG 波形を微分した dEGG 波形のピーク検出によって行なうが、実際の EGG 信号の分析では、多くの未検出、誤検出が観察されることはよく知られており、声門開放時間率の代わりに、閾値を設定した声門開放閾値率を用いて安定して検出をおこなう代替方法が用いられることも多かった。本研究課題の中で、河原-榊原によって、余弦級数包絡の複素 wavelet 分析に基づく簡易な GCI, GOI の検出方法が開発され、閾値設定することなく、精度の高い GCI, GOI の検出が可能となり、EGG 波形からの検出エラーが極めて少ない声門開放時間率の測定が可能となった。

また、榊原の提案による付値-声プロファイルの方法により、基本周波数-発声音圧レベルを点とする二次元空間に、声門開放時間率などの値を付値として色などでマッピングすることにより、個人の声質変化を二次元の中で捉えることが可能となった。声質において、声の基本周波数、緊張度、および発声音圧レベルは強く関連し、声質を規定する、変化させる要因として発声の基本周波数、発声音圧が重要であることが示唆された。

## 5 . 主な発表論文等

[雑誌論文](計 13 件)

1. H. Kawahara, K.-I. Sakakibara, H. Banno, M. Morise, T. Toda, and T. Irino, Aliasing-free implementation of discrete-time glottal source models and their applications to speech synthesis and F0 extractor evaluation, 2015 APSIPA, 520—529, 2015. 査読有, DOI: 10.1109/APSIPA.2015.7415325
2. H. Yokonishi, H. Imagawa, K.-I. Sakakibara, A. Yamauchi, T. Nito, T. Yamasoba, and N. Tayama, Relationship of various open quotient with acoustic property, phonation types, fundamental frequency, and intensity, *J. Voice*, 30(2):145—157, 2016. 査読有, DOI: 10.1016/j.jvoice.2015.01.009
3. A. Yamauchi, H. Yokonishi, H. Imagawa, K.-I. Sakakibara, T. Nito, N. Tayama, and T. Yamasoba, Quantification of vocal fold vibration in various laryngeal disorders using high-speed digital imaging, *J. Voice*, 30(2):205—214, 2016. 査読有, DOI: 10.1016/j.jvoice.2015.04.016
4. A. Yamauchi, H. Yokonishi, H. Imagawa, K.-I. Sakakibara, T. Nito, N. Tayama, and T. Yamasoba, Vocal fold vibration in vocal fold atrophy: quantitative analysis with high-speed Digital Imaging. *J. Voice*. 29(6):755—62, 2015. 査読有, DOI: 10.1016/j.jvoice.2014.12.008
5. 山内彰人, 今川博, 榊原健一, 横西久幸, 大塚満美子, 後藤多嘉緒, 二藤隆春, 田山二郎, 独立成分分析による混合筋電信号の分析方法の研究開発, *音声言語医学* 57(2): 227 -237, 2016, 査読有, <https://doi.org/10.5112/jjlp.57.227>
6. K.A. Stevens, R. Shimamura, H. Imagawa, K.-I. Sakakibara, and I.T. Tokuda, Validation stereo endoscopy with a synthetic vocal fold model, *Acta Acustica united with Acustica*, 102:745—751, 2016. 査読有, DOI: <https://doi.org/10.3813/AAA.918990>
7. A. Yamauchi, H. Yokonishi, K.-I. Sakakibara, T. Nito, N. Tayama, T. Yamasoba, Visualization and estimation of vibratory disturbance in vocal fold scar using high-speed digital imaging, *J. Voice*. 30(4):493—500, 2016. DOI: 10.1016/j.jvoice.2015.07.003
8. A. Yamauchi, H. Yokonishi, H. Imagawa, K.-I. Sakakibara, T. Nito, N. Tayama, and T. Yamasoba, Characterization of vocal fold vibration in sulcus vocalis using high-speed digital imaging., *J. Speech Lang. Hear. Res.*, 60(1):24—37, 2017 査読有, DOI: 10.1044/2016\_JSLHR-S-14-0285
9. H. Kawahara, K.-I. Sakakibara, M. Morise, H. Banno, and T. Toda, A Modulation property of time-frequency derivatives of filtered phase and its application to aperiodicity and  $f_0$  estimation, *Proc. Interspeech 2017*, 424—428, 2017. 査読有, DOI: 10.21437/Interspeech.2017-436
10. H. Kawahara, K.-I. Sakakibara, M. Morise, H. Banno, T. Toda, and T. Irino, A New cosine series antialiasing function and its application to aliasing-free glottal source models for speech and singing synthesis, *Proc. Interspeech 2017*, 1358—1362, 2017. 査読有, DOI: 10.21437/Interspeech.2017-15
11. Y. Li, K.-I. Sakakibara, D. Morikawa, and M. Akagi, Commonalities of glottal sources and vocal

tract shapes among speakers in emotional speech, in *Studies on Speech Production, Lecture Notes in Computer Sci.*, pp. 24–34, Springer. 査読有, 2018. DOI:10.1007/978-3-030-00126-1.

12. H. Kawahara, K.-I. Sakakibara, M.Morise, H. Bannno, T.Toda, and T. Irino, Frequency domain variants of velvet noise and their application to speech processing and synthesis, *Proc. Interspeech 2018*, 2027–2031, 2018. 査読有, DOI: 10.21437/Interspeech.2018-43
13. 河原英紀, 榊原健一, 余弦級数包絡の複素 wavelet 分析に基づく簡易な GCI, GOI の検出方法について, *信学技報*, vol. 118, no. 198, SP2018-22, pp. 1–5, 2018 年 8 月. 査読無, SP2018-222018, Print edition: ISSN 0913-5685 Online edition: ISSN 2432-6380

[学会発表](計 23 件)

1. 榊原健一, D.Sommer, 徳田功, 今川博, 山内彰人, 横西久幸, 田山二郎, ステレオ側視鏡を用いた声帯振動の in vivo 観測による声帯縁の上下変位の計測, 日本音響学会 2015 年秋季研究発表会, 2015 年 9 月, 会津.
2. K.-I. Sakakibara, Interpretation of laryngeal/glottal parameters between physiology, physics, acoustics, and perception, National Conference on Man-Machine Speech Communication, China, Tenjin, 2015 年 10 月, 基調講演
3. H. Kawahara, K.-I. Sakakibara, H. Banno, M. Morise, T. Toda, and T.Irino, Aliasing-free implementation of discrete-time glottal source models and their applications to speech synthesis and F0 extractor evaluation, 2015 APSIPA, 香港, 中国. 2015 年 12 月.
4. K.-I. Sakakibara, H. Imagawa, I.T.Tokuda, A. Yamauchi, H. Yokonishi, and N. Tayama, Analysis of spatial characteristics of the larynx using high-speed digital imaging, ICVPB 2016, 招待講演, Vina del Mar, Chile, 2016 年 3 月.
5. H. Kawahara, SparkNG: Interactive Matlab tools for introduction to speech production, perception and processing fundamentals and application of the aliasing-free-L-F model component, Interspeech 2016, San Francisco, USA, 2016 年 9 月.
6. 山内彰人, 横西久幸, 今川博, 榊原健一, 二藤隆春, 山岨達也, 田山二郎, 声門開閉時間率 (open quotient: OQ) と音質の関連における健常者の性差・年齢差, 日本音声言語医学会学術講演会, 2016 年 11 月, 横浜.
7. 堤内亮博, 山内彰人, 今川博, 榊原健一, 横西久幸, 田山二郎, 喉頭高速度デジタル撮像法によるストロボ画像化に関する研究, 日本音声言語医学会学術講演会, 2016 年 11 月, 横浜.
8. 河原英紀, 榊原健一, エイリアシングの無い声帯音源モデルおよび対話的音声生成シミュレータの拡張について, 日本音響学会 2017 年春季研究発表会, 2017 年 3 月, 神奈川.
9. 李永偉, 榊原健一, 森川大輔, 赤木正人, Commonalities and differences of glottal sources and vocal tract shapes among speakers in emotional speech, 日本音響学会 2017 年春季研究発表会, 2017 年 3 月, 神奈川.
10. H. Kawahara, K.-I. Sakakibara, M.Morise, H.Banno, and T.Toda, A Modulation Property of Time-Frequency Derivatives of Filtered Phase and its Application to Aperiodicity and  $f_0$  Estimation, Interspeech 2017, 2017 年 8 月, Stockholm, Sweden.
11. H. Kawahara, K.-I. Sakakibara, M.Morise, H.Banno, T.Toda, and T. Irino, A New Cosine Series Antialiasing Function and its Application to Aliasing-Free Glottal Source Models for Speech and Singing Synthesis, Interspeech 2017, 2017 年 8 月, Stockholm, Sweden.
12. 李永偉, 榊原健一, 森川大輔, 赤木正人, Relationship between features of glottal sources and vocal tract shapes and perceived positions on valence and activation in emotional speech, 日本音響学会 2017 年秋季研究発表会, 2017 年 9 月, 愛媛.
13. 河原英紀, 榊原健一, 対話的音声生成シミュレータの時変モデルへの拡張について, 日本音響学会 2017 年秋季研究発表会, 愛媛, 2017 年 9 月
14. 河原英紀, 榊原健一, 対数領域パルスにおける声帯音源モデルの拡張について, 日本音響学会 2018 年春季研究発表会, 埼玉, 2018 年 3 月.
15. 若狭健太, 榊原健一, 平賀謙, 寺澤洋子, 声帯振動に着目したオペラ歌唱と合唱歌唱の比較検討, 日本音響学会 2018 年春季研究発表会, 埼玉, 2018 年 3 月.
16. D. Erickson, K.Obert, R. Hayashi, C.Zhu, T. Sadanobu, K.-I. Sakakibara, Is “hanage” really nasal?, 日本音響学会 2018 年春季研究発表会, 埼玉, 2018 年 3 月
17. K. Wakasa, K.-I. Sakakibara, Y. Hiraga, H.Terasawa, The analysis of choral and bel-canto singing with acoustical features and EGG measurement, the Voice Foundation 47<sup>th</sup> annual symposium, Philadelphia, USA, 2018 年 6 月.
18. 河原英紀, 榊原健一, 余弦級数包絡の複素 wavelet 分析に基づく簡易な GCI, GOI の検出方法について, 日本音響学会音声研究会, 京都, 2018 年 8 月
19. H. Kawahara, K.-I. Sakakibara, M.Morise, H. Bannno, T.Toda, and T. Irino, Frequency domain variants of velvet noise and their application to speech processing and synthesis, Interspeech 2018.2018 年 9 月, Hyderabad, India.
20. 若狭健太, 榊原健一, 河原英紀, 寺澤洋子, 生理・音響的特徴量によるオペラ歌唱と合唱歌唱の比較検討, 日本音響学会 2018 年秋季研究発表会, 大分, 2018 年 9 月.
21. 李永偉, 榊原健一, 赤木正人, Simultaneous estimation of glottal source waveform and vocal

- tract shape from speech signal based on ARX-LF model, 日本音響学会 2018 年秋季研究発表会, 大分, 2018 年 9 月 .
22. 若狭健太, 榊原健一, 河原英紀, 寺澤洋子, Open quotient に注目したオペラ歌唱と合唱歌唱の検討, 日本音響学会 2019 年春季研究発表会, 東京, 2019 年 3 月 .
23. 河原英紀, 榊原健一, 森勢将雅, Baseline design of VOCODER based on interference-free spectral representation and generalized excitation source representation, 日本音響学会 2019 年春季研究発表会, 東京, 2019 年 3 月 .

〔図書〕(計 0 件)

〔産業財産権〕

出願状況 (計 0 件)

取得状況 (計 0 件)

〔その他〕

ホームページ等

## 6 . 研究組織

### (1)研究分担者

研究分担者氏名：榊原 健一

ローマ字氏名：SAKAKIBARA, Ken-Ichi

所属研究機関名：北海道医療大学

部局名：リハビリテーション科学部

職名：准教授

研究者番号 (8 桁): 80396168

研究分担者氏名：林 良子

ローマ字氏名：HAYASHI, Ryoko

所属研究機関名：神戸大学

部局名：国際文化科学研究科

職名：教授

研究者番号 (8 桁): 20347785

研究分担者氏名：後藤 多嘉緒

ローマ字氏名：GOTO, Takao

所属研究機関名：東京大学

部局名：医学部附属病院

職名：助教

研究者番号 (8 桁): 20735930

研究分担者氏名：河原 英紀

ローマ字氏名：KAWAHARA, Hideki

所属研究機関名：和歌山大学

部局名：学内共同利用施設等

職名：名誉教授

研究者番号 (8 桁): 40294030

研究分担者氏名：牧 勝弘

ローマ字氏名：MAKI, Katuhiro

所属研究機関名：愛知淑徳大学

部局名：人間情報学部

職名：教授

研究者番号 (8 桁): 50447033

研究分担者氏名：齋藤 毅  
ローマ字氏名：SAITOU, Takeshi  
所属研究機関名：金沢大学  
部局名：電子情報通信学系  
職名：助教  
研究者番号（8桁）：70446962

研究分担者氏名：山川 仁子  
ローマ字氏名：YAMAKAWA, Kimiko  
所属研究機関名：尚絅大学  
部局名：文化言語学部  
職名：准教授  
研究者番号（8桁）：80455196

研究分担者氏名：天野 成昭  
ローマ字氏名：AMANO, Shigeaki  
所属研究機関名：愛知淑徳大学  
部局名：人間情報学部  
職名：教授  
研究者番号（8桁）：90396119

研究分担者氏名：山内 彰人  
ローマ字氏名：YAMAUCHI, Akihito  
所属研究機関名：国立研究開発法人国立国際医療研究センター  
部局名：その他部局  
職名：耳鼻咽喉科医師  
研究者番号（8桁）：90612507

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。