

協和性理論を用いた歌唱時音声の分析手法に関する研究

内田 遼^{1,a)} 矢向 正人^{2,b)}

概要: うなりに基づく不協和を定量化する協和性理論を用いて歌唱時音声进行分析することにより、西洋音楽史における協和・不協和理論の変遷を考察する手がかりを得られる可能性がある。しかし、協和性理論を歌唱時音声に適用するための手法は未整理である。本稿では、歌唱時音声进行分析するために分析手順の改善を行い、歌唱時音声の分析を目指す。

1. はじめに

Helmholtz は、音のうなりに基づく協和・不協和の概念を提唱した [1]。Helmholtz によれば、うなりにより生じるラフネスに由来する音の濁りが不協和である。また、うなりが生じないとき、その音は協和である。

Helmholtz による実験では、刺激音として弦楽器が用いられた。後に、Plomp と Levelt [4] により、純音を用いた実験が行われた。この実験から、Helmholtz による協和・不協和概念を音色一般に拡張するための土台が作られた。Plomp と Levelt の実験により求められた周波数差と不協和の度合いとの関係を表す 2 次曲線（不協和度曲線）から、不協和度を算出し定量化する手法が考案された。この手法は一般に協和性理論と呼ばれる。

協和性理論により算出される不協和度は、うなりが基準となっている。この不協和度は周波数比や音名に基づく協和・不協和とは独立した概念である。協和性理論では、分析対象音のスペクトル構造を用いて計算を行う。したがって協和性理論において協和・不協和特性を左右するのは音色である。音色は、周波数比や音名に基づく協和・不協和概念において無視される概念である。以上のことから、協和性理論を用いて協和・不協和特性进行分析することにより、従来の協和・不協和に関する議論とは異なる観点からの考察が可能になると考えられる。

しかし、協和性理論と音楽に関する先行研究では、不協和度の算出手順が非常に曖昧である。その曖昧さは、協和性理論そのものよりも、協和性理論を用いるための準備段

階において顕著である。したがって本研究では、協和性理論を用いた歌唱時音声の分析の精度を向上させるため、まず分析手法の検討、整理を行う。次に、検討により改善を行った手法を用いて歌唱時音声の分析を行う。

2. 分析手法の検討

歌唱時音声の分析手順を示す。

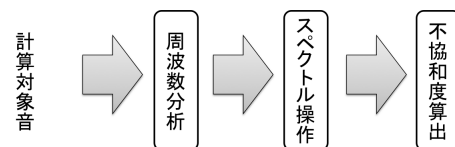


図 1 分析手順の概要

2.1 周波数分析手法

不協和度の算出には計算対象音のスペクトルを用いる。したがって、最初に計算対象音を周波数分析し、スペクトル構造を算出する必要がある。

歌唱法によっては、音声スペクトルが非調波構造となる可能性もある。非調波成分を検出するには、高精度の周波数分析を行う可能性がある。しかし、先行研究 [5][12] では、周波数分析には単なるゼロ詰め FFT が用いられ、厳密な周波数分析が行われたとは言えない。したがって、より精度の高い周波数分析手法の適用が必要である。

協和性理論を用いて周波数差と不協和度とのグラフを作成する際、計算対象音を 2 つ用意した上で、一方の基本周波数を固定し、もう一方の基本周波数を変化させ、それぞれの周波数差について不協和度を計算する必要がある。基本周波数を変化させる際、音域に関係無くスペクトル構造が一定の音であれば、単に基本周波数を変化させるだけでよい。しかし、音声は音域によりスペクトル構造が異なる

¹ 九州大学大学院 芸術工学府
Graduate school of design, Kyushu university, Fukuoka 815-8540, Japan

² 九州大学大学院 芸術工学研究院
Faculty of design, Kyushu university, 同上

a) ryo.uchida.715@s.kyushu-u.ac.jp

b) yako@design.kyushu-u.ac.jp

ため、単に基本周波数を変化させるだけではスペクトル構造が本来の構造と異なってしまふ。したがって、歌唱時音声の分析において、スペクトル構造を実際の音声に対応させながら対象音の基本周波数を変化させる手法が必要である。

周波数分析手法の1つである一般調和解析 (GHA) は、高精度であり、かつ時間分解能と周波数分解能が独立である、また窓関数の影響も受けないとされる [13][14]。したがって、GHA は本研究に最適な周波数分析手法であるように見える。しかし、GHA は本質的に DTFT からのピーク抽出と変わらず、高精度の成分抽出を行うための条件が非常に限定的であることが安部 [8] により報告されている。また、GHA は FFT に比べ計算量が非常に多い。したがって、最大のメリットとされる精度の高さが実現されないのであれば、GHA はきわめて不便な手法であると言わざるをえない。

以上のことより、本研究では QIFFT を周波数分析手法として採用する。QIFFT は、FFT により算出したスペクトルについて、ピークおよびその前後の計3点から真のピークを2次補間により求め、真のピークの周波数、振幅、および位相を抽出する手法である。QIFFT は、パラメータを調整することによりほぼ最尤推定法として機能することが安部および Smith[9] により報告されている。また、QIFFT は FFT を用いた手法であるため、QIFFT を用いることにより GHA に比べ計算の大幅な高速化が期待される。

したがって、QIFFT は高精度および高速計算を兼ね備えた手法であり、本研究で用いる周波数分析手法として最適であると言える。

2次補間には、Smith および Serra[6] による PARSHL アルゴリズムの一部を用いる。精度を上げるため振幅スペクトルを dB 表示に変換した上で、

$$\alpha \triangleq 20 \log_{10} |\tilde{x}'_m(e^{j\omega(k_\beta-1)})| \quad (1)$$

$$\beta \triangleq 20 \log_{10} |\tilde{x}'_m(e^{j\omega k_\beta})| \quad (2)$$

$$\gamma \triangleq 20 \log_{10} |\tilde{x}'_m(e^{j\omega(k_\beta+1)})| \quad (3)$$

(k_β : スペクトルのピークを示す点)

とおく。このとき2次補間による放物線のピークをとる位置と k_β との距離 p は

$$p = \frac{1}{2} \frac{\alpha - \gamma}{\alpha - 2\beta + \gamma} \quad (4)$$

となる。したがって放物線のピーク、すなわち真のピークをとる周波数の推定値は

$$f = (k_\beta + p) \frac{f_s}{N} \quad (5)$$

(N : FFT サイズ)

となり、真のピークの推定値は

$$\hat{y}(p) = \beta - \frac{1}{4}(\alpha - \gamma)p \quad (6)$$

となる。

2次補間を行う際、補間による推定ピーク値と実際のピーク値との間に誤差が生じる。Abe and Smith[9] によれば、ゼロ詰め係数

$$Z_p = \frac{N}{M} \quad (7)$$

(M : 窓サイズ)

を調整することにより、誤差を最小にしつつ計算コストを抑えることができる。ゼロ詰め係数と誤差との関係を図2に示す。

Max Bias	Frequency				Amplitude			
	Rc	Hn	Hm	Bl	Rc	Hn	Hm	Bl
1.00%	2.1	1.2	1.2	1.0	2.6	1.4	1.4	1.1
0.50%	2.5	1.5	1.5	1.1	3.0	1.6	1.7	1.2
0.10%	4.1	2.4	2.4	1.9	4.3	2.3	2.4	1.8
0.01%	8.7	5.0	5.1	4.0	7.5	4.0	4.2	3.1

Rc: Rectangular, Hn: Hann, Hm: Hamming, Bl: Blackman.

図2 所定誤差バイアス以下に抑えるために必要な最小のゼロ詰め係数 ([9] より)

また、FFT 点数は FFT 結果の周波数間隔に影響を与える。周波数間隔と誤差との関係を図3に示す。

window	Theory	Experiment					
		Z_p	MAFS	MAFS	Bias (%)		
						Freq	Amp
Rect		2.0	1.94	1.90	15.9	24.6	5.37
		3.5	1.73	1.39	16.3	23.2	7.14
		5.0	1.64	1.38	16.5	22.1	6.95
Hann		2.0	2.87	2.38	3.89	4.34	1.26
		3.5	2.66	2.30	4.09	2.88	0.91
		5.0	2.57	2.28	4.15	2.74	0.87
Hamm		2.0	2.72	2.35	1.37	2.19	0.60
		3.5	2.51	2.20	1.42	0.87	0.27
		5.0	2.42	2.18	1.45	0.76	0.24
Black		2.0	3.53	3.23	0.39	0.28	0.07
		3.5	3.32	3.05	0.39	0.13	0.04
		5.0	3.23	3.00	0.40	0.13	0.04

図3 最小許容周波数間隔と最大誤差バイアス ([9] より)

図2および図3より、窓の種類によって誤差と対応するゼロ詰め係数と周波数間隔の値が異なる。本稿では、ゼロ詰め係数をなるべく低く抑えた上でピーク推定精度を上げるため、Blackman 窓を用いる。窓サイズは4096点、FFT 点数は16384点に設定する。ゼロ詰め係数は4.0となり、図2において誤差バイアスは0.01%以下に抑えられる。また、図3において誤差バイアスは周波数: 約0.39%、振幅: 約0.13%、位相: 約0.04%となる。たとえば周波数の誤差バイアスについて、0.39%をセント値換算すると約0.56centとなる。また、振幅の誤差バイアスについて、0.13%をdB

に換算すると約 0.011dB の差となる。後述する協和性理論のアルゴリズムにおいて、これらの値は計算に実質的な誤差をもたらさない。したがって、本研究における QIFFT の使用は妥当であると言える。

同一の音声データを通常の FFT, QIFFT で分析した結果を図 4, 5 に示す。また、不協和度を算出するため、QIFFT の結果からピーク抽出を行った。図 5 の結果から次数の低い順にピークを 22 本抽出した結果を図 6 に示す。

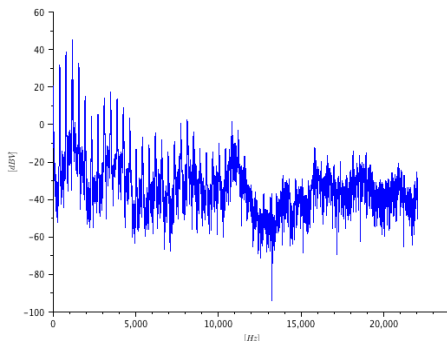


図 4 通常の FFT の結果

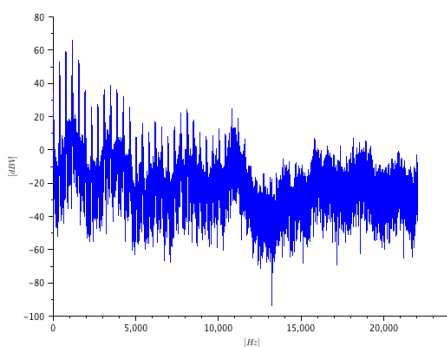


図 5 QIFFT の結果

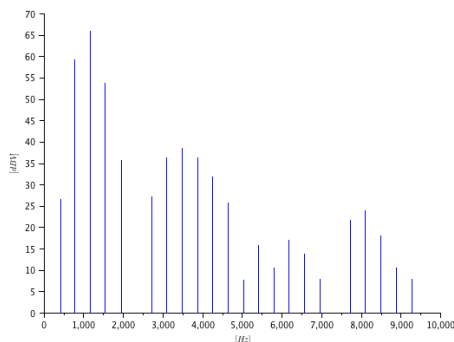


図 6 QIFFT の結果, ピーク抽出

2.2 スペクトル構造近似手法

本稿では、まず分析対象音のフォルマント周波数を推定する。そして得られたフォルマント周波数をカットオフ周波数とみなし、個々のスペクトルのレベルを近似する。

上記の手順を行うために、フォルマント周波数の推定が必要となる。本稿では、フォルマント推定手法として線形予測分析を採用する。2.3 で周波数分析に用いたデータの最初から 200 点を用い、線形予測分析を行う。フィルタ係数の次数は 32 とする。同一の歌手が 5 つの音域を歌い分けたサンプルについて線形予測分析を行った例を、図 4 に示す。

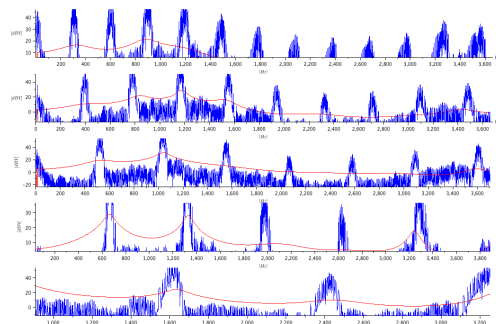


図 7 線形予測分析によるフォルマント推定の例

推定されたフォルマント周波数からカットオフ周波数を設定する際、幾つのフォルマントをカットオフ周波数とみなすかが問題となる。歌唱時音声、特に男声および低 - 中音域の女声によるオペラの歌唱法については、スペクトル構造の中でも歌唱フォルマントが特徴的である。したがって、少なくとも歌唱フォルマントが生じる音声については、低次から歌唱フォルマントが生じる 2 - 3KHz 周辺まで計算対象音を抽出する必要がある。

2.3 協和性理論の選択

不協和度算出手法である協和性理論は、複数存在する。代表的な手法として、Sethares の手法 [5]、亀岡・厨川の手法 [2][3][11]、亀岡・厨川の手法の拡張である小畑の手法 [12] が挙げられる。それぞれの手法について、不協和度の算出手順は共通である。

- (1) 分析対象音を 2 つ用意する (a, b とする)。
- (2) a と b が同時に鳴っていると仮定し、a および b のすべての成分音のペアについて、不協和度を算出する。
- (3) 分析した全てのペアの不協和度を加算し、a および b が同時に鳴る場合の全体の不協和度とする。

Sethares の手法では、成分音の算出に Plomp および Lev-elt による不協和度曲線が用いられる。一方、亀岡・厨川の手法では、亀岡・厨川が独自に実験を行い導出した不協和度曲線が用いられる。

Sethares は、純音のペアの不協和度が 0 から 1 までのオーダーであり、不協和度を定性的な量としてみなす。ま

た、加算の際に、不協和度を単純加算している。一方、亀岡・厨川は、不協和度を心理量とみなし、Stevensのべき乗則を用いて物理量に相当する不協和強度へと変換された上で加算することより、不協和度を定量的に扱い計算を行うためには、亀岡・厨川の考え方を採用することが妥当だと言える。

小畑は、亀岡・厨川の手法の改善を試みている。亀岡・厨川の手法では、計算対象とする成分音のペアが近い周波数を持っていても、音の合成は行われなかった。周波数が近い成分音のペアは高い不協和度を示しやすいため、亀岡・厨川の手法を用いたのでは計算結果が不自然に高くなる可能性が考えられる。小畑はこの問題を解決するため、2つの成分音の周波数 $f_1, f_2 (f_1 < f_2)$ について

$$1.01f_1 > f_2 \quad (8)$$

となる場合、2つの音が合成されるとみなした。成分音の振幅をそれぞれ d_1, d_2 とおくと、合成後の周波数 f 、振幅 B はそれぞれ

$$f = \frac{f_1 + f_2}{2} \quad (9)$$

$$B = \sqrt{d_1^2 + d_2^2} \quad (10)$$

となる。

本稿では、3つの協和性理論の内、小畑の手法を採用することにしている。亀岡・厨川の実験 [3] により、不協和度を心理量とみなすことにより、実験値と理論値がよく一致することがわかっている。それに加え、小畑の手法は音の合成を考慮に入れている点で、亀岡・厨川の手法の欠点を補っていると言える。したがって、小畑の手法の採用は妥当だと言える。

3. 歌唱時音声の分析に向けて

2章で検討した一連の手順を実装中である。実装には scilab を用いている。本稿執筆現在、QIFFT および小畑の手法の実装は完了している。フォルマント推定から計算対象成分音を近似し抽出するアルゴリズムを、現在作成中である。

4. 今後の展望

4.1 西洋クラシック音楽における協和・不協和理論の変遷の考察

西洋クラシック音楽において、協和・不協和の定義および理論は時代と共に変化してきた。ポリフォニー以前の時代は周波数比が単純な整数比であることが協和であるとされ、ポリフォニーの成立と共に音名に基づく協和・不協和が普及した。中世からルネサンス期にかけて、それまで協和音程とみなされていた完全4度は不協和音程とみなされるようになった。逆に、それまで不協和音程とみなされて

いた長短の3度は、協和音程とみなされるようになった。この逆転現象は、はイギリス・アイルランド地方の純正3度音程を好む音楽スタイルが大陸に伝来し、フランスにおいて3度や6度を多用するフォーブドロンが生まれたことによるとされる [10]。しかし、従来の協和・不協和概念が完全に逆転してしまうという特異な現象は、単なる理論的側面から説明できるとは考えにくい。新しい協和音程が普及するためには、その音程が耳に馴染む過程が必要である。ここで、歌唱時音声を用いて分析することにより、各音程と不協和度との関係を明らかにすることができる。当然、完全4度と長短3度との不協和度の大小関係も算出できる。したがって、長短3度が人々の耳に馴染んだ過程を、不協和度の観点から説明できる可能性がある。

5. おわりに

協和性理論を用いた歌唱時音声の分析手法の改善を試みてきた。本稿で選択したアルゴリズムを実装中である。分析の精度の高さも必要であるが、アプリケーションへの適用等も考えれば、アルゴリズムの高速化も必要になる。今後の課題として、引き続き実装および検討を続ける。

参考文献

- [1] Helmholtz, H.L.F.: *Die Lehre von den Tonempfindungen als physiologische Grundlage für die Theorie der Musik*, F. Vieweg (1870) (*On the Sensations of Tone as a Physiological Basis for the Theory of Music*, with a new introduction by Henry Margenau, Dover (1954)).
- [2] Kameoka, A., Kuriyagawa, M.: Consonance Theory Part I: Consonance of Dyads, *Journal of the Acoustical Society of America*, Vol. 45, No. 6, pp. 1452 - 1459 (1969).
- [3] Kameoka, A., Kuriyagawa, M.: Consonance Theory Part II: Consonance of Complex Tones and Its Calculation Method, *Journal of the Acoustical Society of America*, Vol. 45, No. 6, pp. 1460 - 1469 (1969).
- [4] Plomp, R., Levelt, W.J.M.: Tonal Consonance and Critical Bandwidth, *Journal of the Acoustical Society of America*, Vol. 38, pp. 548-560 (1965).
- [5] Sethares, W.A.: *Tuning, Timbre, Spectrum, Scale*, Springer (1998).
- [6] Smith III, J.O., Serra, X.: PARSHL: An Analysis / Synthesis Program for Non-Harmonic Sounds Based on a Sinusoidal Representation, International Computer Music Conference, STAN-M-43 (1987).
- [7] Sundberg, J.: *The Science of the Singing Voice*, Northern Illinois University Press (1987) (榎原健一監訳、伊藤みか、小西知子、林良子訳: 歌声の科学、東京電機大学出版局 (2007)).
- [8] 安部素嗣: 一般化調和解析の周波数領域での振舞い - 本当にGHAはFFTより精度がよいのか?-, 電子情報通信学会技術研究報告. EA, 応用音響 105(136), pp. 13-18 (2005).
- [9] 安部素嗣, Smith, J.O.: FFTの2次補間に基づく正弦波パラメータ推定法の設計基準 ~擬似定常な正弦波成分の場合~, 電子情報通信学会技術研究報告. EA, 応用音響 104(304), pp. 7-12(2004).
- [10] 藤枝守: 響きの考古学 音律の世界史, 音楽之友社 (1998).
- [11] 亀岡秋男: 協和性理論とその応用, 慶応義塾大学大学院博士号取得論文 (1974).

- [12] 小畑郁男: 楽器の音色を視野に入れた音高構成理論の研究 -感覚的協和理論の音楽への応用-, 芸術工学博士 (九州芸術工科大学) 論文 (2002).
- [13] 牛山聡, 東山三樹夫, 飯塚昌弘, 平田能睦: 一般調和解析による波形分析, 電子情報通信学会技術研究報告. EA, 応用音響 93(527), pp. 39-44 (1994).
- [14] 山崎芳男: 音響信号の時間周波数分析, 日本音響学会誌 53 巻 2 号, pp. 147-153 (1997).