

LE-004

# 単母音による歌唱音声スペクトルの統計的分析に基づく音色制御法の提案と評価 Timbre control of singing voice based on statistical analysis of singing vowel spectra and its evaluation

森勢 将雅†  
Masanori Morise

田原 佳代子†  
Kayoko Tahara

高橋 徹†  
Toru Takahashi

入野 俊夫†  
Toshio Irino

河原 英紀†  
Hideki Kawahara

## 1. まえがき

近年、多様な音声合成を実現するための研究が進められている。音声分析変換合成法 STRAIGHT[1,2]を用いることで、音量・音色を独立して操作しつつ、入力音声に匹敵する品質の音声を合成できる。STRAIGHT の提案により、従来の技術では困難であった歌唱表現とスペクトルとの関係を、精密に分析できるようになった。そのため、近年では自然な歌唱を合成するための技術の提案[3,4,5]に用いられている。本研究は、人間が自然に感じる歌声の合成を目標としている。スペクトルや基本周波数等、物理的なパラメータを制御することでこれら主観的な印象を制御することが出来れば、自然で豊かな歌唱合成だけでなく音声合成の自然性向上や感情制御など様々な応用が期待できる。

これまでの研究では、RWC 研究用音楽データベース[6]に含まれる歌手の音声を素材とした合唱曲を合成した。この合唱曲は、RENCON'04 の聴き比べにおいて最も品質の良い合成歌唱システムであると評価された[7]。この合唱曲の歌唱合成では、素材とした歌唱音声の時間長、音量、音高のみを制御し、音量や音高の変化に伴うスペクトルの制御を行っていない。歌唱音声の音量や音高に対応するスペクトル変化に共通する成分が認められるならば、音量や音高の制御に加え対応するスペクトル成分を付与することで、より自然な歌唱音声を合成できる可能性がある。

本論文では、RWC 研究用音楽データベースに含まれる単母音から音量の変化に相当するスペクトル変化を抽出し、歌唱音声合成へ組み込む方法を示す。この歌唱合成法は、音量を変化させる歌唱法であるクレッシェンド・デクレッシェンド歌唱をより自然に合成できることが期待できる。図 1 に提案する音色制御法の概略を示す。RWC 研究用音楽データベースに収録されている単母音歌唱音声を STRAIGHT で分析し、音量の変化に伴うスペクトル変化を主成分分析により求める。求めた主成分から音量 1 dB の変化に対応するスペクトルを求め、入力素材となる音声と音量とを用いた歌唱合成法を示す。提案する手法を用いて合成したクレッシェンド歌唱の品質を主観評価し、提案法で合成した歌唱音声は音量のみを制御する方法より自然であることを示す。

## 2. 歌唱音声のスペクトル分析

### 2.1 音声素材

本研究では RWC 研究用音楽データベースに収録されている 18 名の歌手(ソプラノ 3 名, アルト 3 名, テノール 3 名, バリトン 3 名, バス 3 名, R&B 系男性 1 名, 女性 2

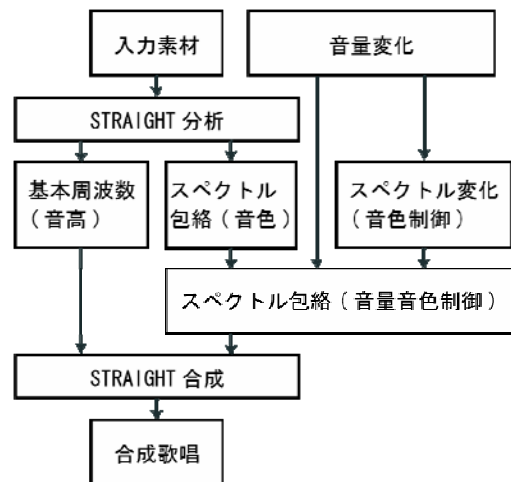


図 1. 提案する音色制御法の概略

名)のノーマル歌唱音声(データベースにはノーマル・スタッカート・ビブラートが収録されている)を分析に用いた。なお、強弱に関しては、個々の奏法、音高について強・中・弱の 3 段階の強さが収録されている。音量、音高、母音毎にデータを区切り、計 4960 個のノーマル歌唱音声を STRAIGHT により分析した。

### 2.2 STRAIGHT を用いたスペクトル分析

STRAIGHT は、音声を基本周波数、スペクトル包絡(以下では STRAIGHT スペクトルとする)、非周期性指標に分解する。STRAIGHT により抽出されたこれらのパラメータを用いて再合成した音声は、元音声に匹敵する品質を持つ。

本論文では、音量の変化に対応するスペクトル変化を調べるため、STRAIGHT スペクトルの主成分分析を行う。歌声は歌い出しと歌い終わりについては特別な処理を行うこととし、開始・終了点の 125 ms を除き 1ms 毎に STRAIGHT スペクトルを得た。

さらに聴覚的印象に与える影響を調べることを狙い、STRAIGHT スペクトルを聴覚抹消系の分析を近似する 1/3 オクターブバンド毎のレベルに変換した。分析に用いる中心周波数  $f_c$  は 50 Hz から 16000 Hz の 26 次元とした。ある中心周波数  $f_c$  における帯域毎のレベル  $L(f_c)$  は、信号  $x(t)$  の STRAIGHT スペクトル  $F(\omega)$  から以下の式により求める。帯域のレベル  $L(f_c)$  の単位は dB とする。

†和歌山大学システム工学研究科

$$L(f_c) = 10 \log_{10} \left( \int_{\omega_L}^{\omega_H} |F(\omega)|^2 d\omega \right)$$

where  $\omega_L = 2\pi f_L$ ,  $\omega_H = 2\pi f_H$  (1)

$$f_L = 2^{-\frac{1}{6}} f_c, f_H = 2^{\frac{1}{6}} f_c$$

なお、STRAIGHT スペクトル  $F(\omega)$  のエネルギーは全てのフレームにおいて等しくなるように正規化した。全ての単母音について式 1 により 26 次元ベクトルの系列を求めた。

歌手、母音と奏法については固定し、音量・音高を変化させた単母音の STRAIGHT スペクトルを全て結合し、スペクトログラム  $S$  とする。あるフレームにおける 26 次元の行ベクトルを  $s_t$  とし、結合したスペクトログラム  $S$  を式 2 により得る。

$$\begin{aligned} S &= [s_1, s_2, \dots, s_N] \\ \bar{S} &= [\bar{s}_1, \bar{s}_2, \dots, \bar{s}_N] \\ \bar{s} &= \frac{1}{N} \sum_{t=1}^N s_t \end{aligned} \quad (2)$$

また、全ての行が  $s_t$  の全フレームについての平均値  $\bar{s}$  とする、スペクトログラム  $S$  と同じサイズの行列を  $\bar{S}$  とする。

結合したスペクトログラム  $S$  の共分散行列を求め、この共分散行列の固有値と、固有値に対応する固有ベクトルを求めた。以下では固有ベクトルを列ベクトルとする正方行列を  $V$  とする。これらの固有値の寄与率は全体の分散をどの程度説明しているか示しており、固有ベクトルは各周波数帯域の重みを示している。このとき得られる寄与率の高い固有ベクトルが、音量の変化に対応するスペクトル変化である。

歌手 18 名、5 母音全てを平均した第 4 主成分までの累積寄与率は 81% であった。図 2 にソプラノ歌手 1 名における第 4 主成分までの固有ベクトルを示す。こうして得られた固有ベクトルから、スペクトルの主要な変動を求める。

### 2.3 音量変化に伴うスペクトル変化

スペクトルの主要な変動は、主成分得点  $p_1, \dots, p_M$  ( $M = 26$ ) から調べることができる。主成分得点  $p_1, \dots, p_M$  を行ベクトルとし、第 26 主成分得点までを列ベクトルとして並べた正方行列を  $P$  とし、以下の式により求める。

$$P = (S - \bar{S})V \quad (3)$$

式 3 により得られた主成分を手がかりとして、音量変化に対応するスペクトル変化の性質を調べた。分析した音声の音圧レベルを独立変数、得られた第 1 主成分得点を従属変数として回帰分析を行った。図 3 にソプラノ歌手 1 名の

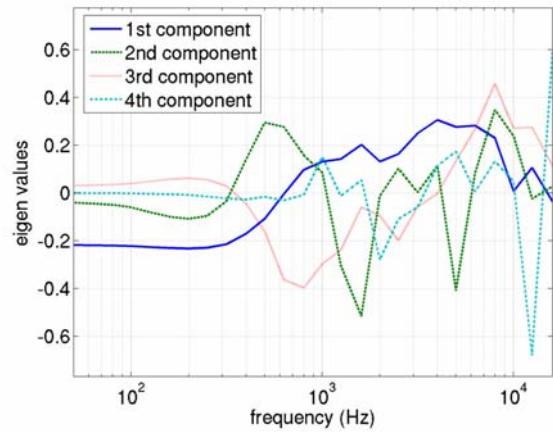


図 2. ソプラノ歌手 1 名における第 1 から第 4 固有ベクトル

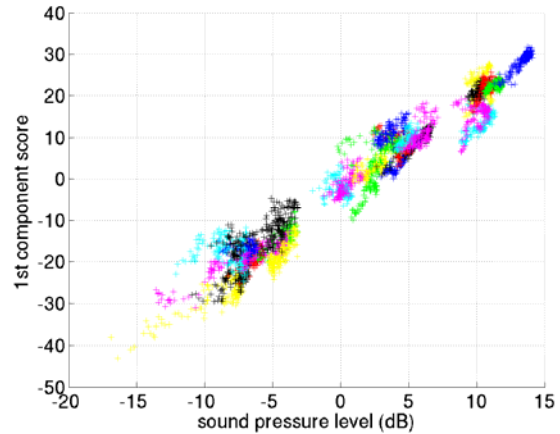


図 3. ソプラノ歌手 1 名の第 1 主成分得点と音量との散布図

第 1 主成分得点と音量との散布図を示す。横軸は音量、縦軸は第 1 主成分得点を示している。主成分分析を行う際「強弱」と「音高」を変化させた単発音声を結合しているため、単音毎に色を変えて表示した。図 3 より、第 1 主成分得点は音量と強い相関があることが分かる。この傾向は 18 名の歌手全てにおいて認められた。そのため、第 1 固有ベクトルは音量の変化に対応するスペクトル変化と考えられる。図 2 の第 1 固有ベクトルより、歌唱音声の音量が増加すると相対的に低域のエネルギーが小さくなり、1 kHz から 8 kHz 程度の帯域を強調していることが分かる。これは、歌声において観測される歌声ホルマント[8]を強調している可能性を示している。

## 3. 歌唱音声合成法への応用

### 3.1 音量変化に伴うスペクトル変化

主成分分析により得られた固有ベクトルを歌唱合成に利用するため、音量 1 dB に相当するスペクトルの変化量へ変換する。本論文では、音量 1 dB あたりのスペクトル変

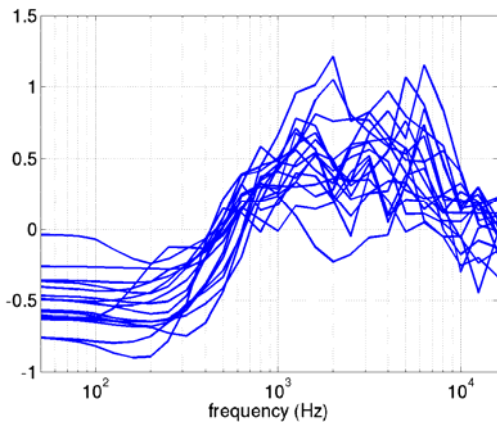


図 4. 18 名の歌手全ての音色制御基底ベクトル

化に相当するよう変換した固有ベクトルを音色制御基底ベクトルと定義する。音色制御基底ベクトルは第 1 主成分得点を従属変数、音量を独立変数として回帰係数を求め、第 1 固有ベクトルに掛けたものとした。なお、歌唱音声の音量は相対的なものとし、単位は dB とした。そのため、歌唱合成の素材として用いる音声の絶対音量ではなく、操作量に依存してスペクトルの変形を行う。

### 3.2 歌手毎の音色制御基底ベクトル

素材となる歌唱音声と音量の情報とを用いて歌唱の合成を行う。図 1 に示すように、音量と音色制御基底ベクトルからスペクトル変化量を定め、STRAIGHT スペクトルに加えることで音色の制御を行う。音色制御基底ベクトルは歌手や母音毎に求めることができる。図 4 に 18 名の歌手全ての音色制御基底ベクトル、図 5 に、5 母音それぞれの音色制御基底ベクトルを示す。歌手毎の音色制御基底ベクトルは、特に高域において広く分布している。しかし、母音による違いは小さい。そのため、歌唱音声合成に用いる音色制御基底ベクトルは、歌手毎に作れば良いと考えられる。

## 4. 提案した音色制御法の評価

### 4.1 音声刺激

提案した音色制御法を用いてクレッシェンド音声の合成を行い、自然性の評価を行った。音声素材は 6 名の歌手による 5 母音の歌唱音声とした。実験に用いた歌手は、ソプラノ、アルト、テノール、バス、R&B(男性)、R&B(女性)各 1 名の計 6 名である。各歌手、各母音に対して以下の 3 条件で合成音声を作成した。

- (1) 音量を制御し、歌手本人から求めた音色制御基底ベクトルを用いてスペクトル制御を行った合成歌唱
- (2) 音量を制御し、他の歌手から求めた音色制御基底ベクトルを用いてスペクトル制御を行った合成歌唱
- (3) 音量のみを制御した合成歌唱

ここでは、全ての条件において音高を固定したクレッシェンド歌唱を合成した。クレッシェンド歌唱のダイナミックレンジは 15 dB とした。なお、条件(2)で使用する他の歌

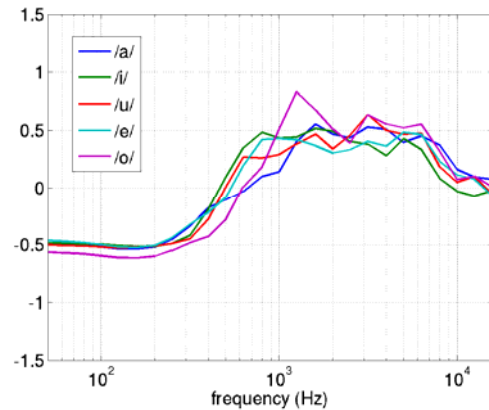


図 5. 日本語 5 母音全ての音色制御基底ベクトル。全ての歌手について平均した。

手の音色制御基底ベクトルは、歌手本人の音色制御基底ベクトルとその他 17 名の歌手の音色制御基底ベクトルから各ユークリッド距離を求め、最大となる歌手の音色制御基底ベクトルとした。この検討により、音色制御基底ベクトルの汎用性について示す。

### 4.2 聴取実験

歌手 1 名 1 母音について作成した 3 つの合成歌唱音声を総当たりとなるように 2 つの歌唱音声をペアとし、提示順序の影響を考慮し 6 組のペアを作成した。聴取実験では 1 条件につき 2 回提示することとした。すなわち、歌手 6 名、5 母音では合計 360 セットの音声を提示する。被験者は正常な聴力を有する大学生・大学院生の男女 8 名とした。1 セットにつき 2 つの音声が提示され、被験者は「どちらが自然か」を回答した。「自然」の判断基準は、実際に歌手がクレッシェンドで声を大きくしたときの音色変化と同様に聴こえる、あるいは実際の歌手の演奏と同様な音色の変化と感じられる場合と教示した。

### 4.3 実験結果

被験者の回答にはばらつきがあったので、各被験者の歌手・母音毎に選択度数の平均値を求め、クラスタ分析を用いて 3 種類に分類した。分類 1 は 2 名、分類 2 は 4 名、分類 3 は 2 名である。各分類について危険率 10% で検定を行った。結果を表 1 に示す。表 1 は、各条件について有意な差が存在するかを示している。例えば(1)>(3)の場合、条件(1)が条件(3)より自然と感ずることに有意差があったことを示している。空欄は、有意差が存在しなかったことを示している。

まず、提案する音色制御基底ベクトルを用いて歌唱音声を合成した条件(1)、(2)と、音量の操作のみを行った条件(3)とを比較する。本人から求めた音色制御基底ベクトルを用いて合成した歌唱と音量のみ制御した歌唱とを比較すると、分類 1 の被験者では、アルト奏者とバス奏者を除く全ての条件で提案法が自然であると解答している。一方、分類 2 と分類 3 の被験者では、ほとんどの条件で有意差が存在せず、どちらが自然であるか明確な傾向は認められない。

表 1. どちらが自然か, についての検定結果

	(1)vs(3)			(2)vs(3)			(1)vs(2)		
	分類 1	分類 2	分類 3	分類 1	分類 2	分類 3	分類 1	分類 2	分類 3
ソプラノ	(1)>(3)	(3)>(1)	(3)>(1)				(1)>(2)	(2)>(1)	
アルト		(1)>(3)		(2)>(3)			(2)>(1)	(1)>(2)	
テノール	(1)>(3)						(1)>(2)		
バス				(2)>(3)	(3)>(2)	(2)>(3)		(1)>(2)	
R&B(男)	(1)>(3)			(2)>(3)	(3)>(2)			(1)>(2)	
R&B(女)	(1)>(3)						(1)>(2)		
母音/a/	(1)>(3)			(2)>(3)	(3)>(2)				(2)>(1)
母音/i/	(1)>(3)			(2)>(3)	(3)>(2)				
母音/u/	(1)>(3)								
母音/e/	(1)>(3)			(2)>(3)					
母音/o/	(1)>(3)			(2)>(3)					
全て	(1)>(3)			(2)>(3)	(3)>(2)	(2)>(3)			

他人の音色制御基底ベクトルを用いて合成した場合と音量のみ制御した場合とを比較すると, 分類 1 の被験者では多くの条件で音色制御基底ベクトルに基づいた合成歌唱のほうが自然であると解答している. 分類 2, 分類 3 の被験者では, 音量のみを制御する条件(3)のほうが自然であると解答している. これらのことから, 被験者により自然性の好みに差があるものの, 本人の音色制御基底ベクトルを用いた場合, 自然性の向上が認められていると考えられる. 本人の音色制御基底ベクトルを用いた条件(1)と, 他人の音色制御基底ベクトルを用いた条件(2)とを比較すると, 全ての被験者において, どちらが自然に聴こえるか明確な傾向が認められていない. これは, 別の歌手のクレッシェンドの特徴を付与しても自然性が損なわれない可能性を示している.

## 5. まとめ

本論文では RWC 研究用音楽データベースに収録されている約 5000 個の単母音歌唱音声を用い, 音量の変化に相当する成分を抽出した. 抽出した成分から音量 1 dB の変化に相当する音色制御基底ベクトルを求め, 歌唱音声合成へ組み込む方法を提案した. 音色制御基底ベクトルを用いて合成したクレッシェンド歌唱を主観評価し, 提案法は母音による差が少ないこと, 被験者により音色制御による自然性の向上が認められることを明らかにした.

今後は, 様々な歌唱音声に対して様々な音色制御基底ベクトルを求め, 歌唱の自然性と音色制御基底ベクトルとの関連を明らかにする.

## 謝辞

本研究は CrestMuse プロジェクト, 科学研究費補助金(特別研究員 18-2150)による支援を受けて行われた.

## 参考文献

- [1] 河原英紀, "高品質音声分析変換合成法 STRAIGHT の出自," 信学技報, SP2005-141, pp.13-18, 2006.
- [2] Hideki Kawahara, Ikuyo Masuda-Katsuse and Alain de Cheveingne, "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds," Speech Communication, vol. 27, no. 3-4, pp. 187-207, 1999.
- [3] 河原英紀, 片寄晴弘, "高品質音声分析変換合成システム STRAIGHT を用いたスキャット生成研究の提案," 情報処理学会論文誌, vol.43, no.2, pp.208-218, 2002.
- [4] Hideki Kawahara, Haruhiro Katayose, Alain de Cheveingne, Roy D. Patterson, "Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation," Proc. 2003 IEEE International conference on acoustics, speech and signal processing(ICASSP2003), vol.1, pp.256-259, 2003.
- [5] 豊田健一, 片寄晴弘, 河原英紀, "STRAIGHT による歌声モーフィングの初期的検討," 情報処理学会研究報告, 第 2006-MUS-64, 2006.
- [6] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一, "RWC 研究用音楽データベース: 音楽ジャンルデータベースと楽器音データベース," 日本音響学会講演論文集, pp.843-844, 2003.
- [7] Hideki Kawahara, Hideki Banno, Masanori Morise, Yumi Hirachi, "A cappella synthesis demonstrations using RWC music database" Proc, NIME04, pp.130-131, 2004.
- [8] Ingo R. Titze, "音声生成の科学 発声とその障害," 医歯薬出版株式会社, pp.143-192, 2003.