バランス歌唱コーパスを用いた
歌唱音高コントロールの統計的分析

Statistical Analysis of Singing Pitch Control using a balanced Singing Voice Corpus

深澤　友貴
Yuki Fukasawa
法政大学情報科学部ディジタルメディア学科
E-mail: yuki.fukasawa.2n@cis.hosei.ac.jp

Abstract

Individual differences appear in singing, which most familiar musical music representation. And if gaps appear in each musical sensitivity, gaps appear in the pitch too; search results greatly fluctuate in search using pitch information. The objective of this research, which focused attention on pitch fault and anteroposterior relationship from false results, is an estimation method for more exact melody by using singing pitch control model. At the first of this study, by selecting some music piece from Singing practice pieces, a singing corpus was structured. It has various alteration, two kinds of beat, tempo and so on. That system estimate the pitch automatically from singing sample, and the results are analyzed statistically averages and standard deviation of each notes. As a result, it has been understood that some individual halftone changes in singing differ from each other and that is effective in statistical modeling. Additionally, maximum accuracy rate was 100% in likelihood estimation for relative singing pitch variation.

1. まえがき

歌唱とは，ある一定の規則に則って並べられた音を歌う事である．発声出来れば誰でも音楽を奏でられるため，人間にとって最も身近な音楽表現である．この歌唱の音高コントロールは個人に依り，歌唱音高にはずれが生じる場合がある．自動採譜システムでは，このずれが結果を左右するが，従来，これを明示的にモデル化したものはなく，また相対的な音高比率や値が変わらない事が前提とされていた[1][2]．ずれによる歌唱者からみた誤推定を解決するには，誤って歌唱された旋律から歌唱者が意図した旋律の推定が考えられる．

本研究では統計的な歌唱モデルを作成し，歌唱旋律の相対音高差にずれがあるものから，歌唱者が意図した相対音高差の並びの推定を目標としている．ある一定の歌唱曲を収録し，その中での相対音高差についてのずれがある程度の分布を示すと考える事により，入力された歌唱音声における音高差から尤度によって，楽譜上での対応する相対音高差の推定が出来る．この研究の応用には，相対音高差からの具体的な階名推定及び旋律の推定，並びに相対音高差情報を検索キーに用いた楽曲検索などが考えられる．

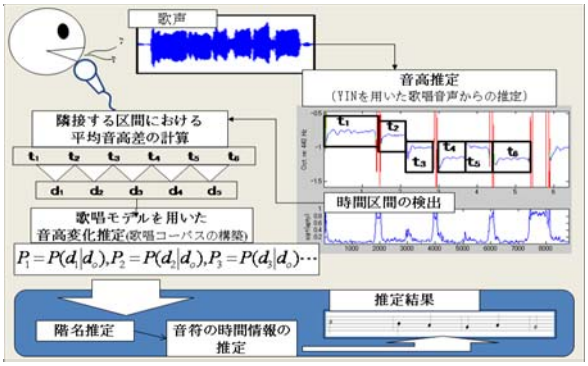

図 1　歌唱からの旋律推定法.

本論文ではまず音高コントロールについての調査を行い，個人における統計的な分布を明らかにした．そしてその分布を用いた相対的な音高差の推定までを行った．

2. 歌唱からの旋律推定法

歌唱からの旋律推定についての研究には清水らによるものがある[2]．これは平均律音階における周波数比と等しい間隔のテンプレートを用いたマッチングにより音階位置を認識，推定された階名から旋律の推定をするものである．しかし相対的にみて正確に歌えている事を条件としているため，比率にずれが生じると誤った推定が行われてしまう．この各階名の周波数比が保たれない事に起因する問題解決のため，歌唱の統計的モデルを用いた旋律推定法を提案する(図 1).

1.歌唱を収録
　wav データ(サンプリング周波数(SR):44.1kHz)
2.波形データから音高を推定
　YIN により推定，C4 を基準に cent 値を計算.
　計算窓長:SR/30 点,シフト:32/SR 点,
3.時間区間の検出
　YIN からの出力(音高情報の非周期性の割合)を用いて区間を検出，音高値の切り分け.
4.隣接する一定音高区間における平均音高の差の計算
　Cent 値で計算.
5.歌唱モデルから相対音高変化を推定

Supervisor: Prof. Katsunobu Ito, Prof. Hiroshi Hanaizumi

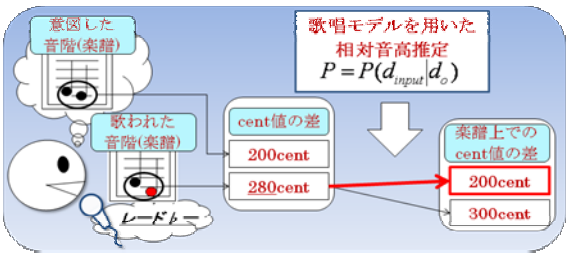

図２　歌唱からの旋律推定法

なお，本研究では歌唱モデルからの音高変化推定方法に着目しているため，階名推定や音符の時間情報の推定については触れていないが，次のような処理が考えられる．

6.階名推定
　推定された相対音高変化を持つ，前後の階名を推定．

7.音符の時間情報の推定
　検出された時間区間から音符長を推定．

8.推定結果の提示
　推定された音階名と音符長との組み合わせとして提示．

2.1. 音高のずれが生じる要因

まず，先に述べたような音高のずれが生じる原因について次のようにそれぞれ仮定した．

1,歌唱者の中にあるイメージ(音階)のずれ
　音感が基準とするものを歌唱者が自身で持つ音階と考えたとき，これと平均律[4]とのずれが表れる．

2,前の音と目標音の関係
　ある音からある音に変化させた時に，個人によって歌い易いものとそうでないものがある．

3,曲の持つ特徴(調，テンポ，強弱など)
　曲調によって歌い易さに違いがある．

4,歌詞の音韻
　声の高さは声帯の振動周期であり，音韻は声道や鼻腔によって決定される[3]．三つの間に直接の関係はないが，音韻によって音高が安定しない事がある．

5,歌唱者の技量
　訓練の量，有無などにより技量が異なる．

これら「歌唱者の音高コントロール」を用い，歌唱音声からの旋律推定を考える．1 の音高のずれは，2 の前後関係を加味することによって確率的に見ることが出来る．この確率を求めるため歌唱コーパスを作成し，音符の前後関係とずれを用いた歌唱の統計的モデルを考える．

2.2. 歌唱モデル

このモデルは(1)式で表わされ，観測値(音高)から相対音高変化を推定するものである(図 2).

$$D=\arg\max_i\{P(d_{input}|d_i)\}\quad(1\le i\le 15)\qquad(1)$$

d_{input} …入力歌唱での相対音高差

d_i …コーパスにおける楽譜上の相対音高差

まず楽譜を用いた歌唱サンプルを収録し，音高差の分布と楽譜情報における音高変化の対応を取った歌唱コーパスを作る．歌唱中の音高情報 S とそれぞれの差を

$$S=t_1,t_2,...,t_n \text{ (t は一定音高区間の平均値)}\qquad(2)$$

$$d_1=t_2-t_1,d_2=t_3-t_2...,d_n=t_n-t_{n-1}\qquad(3)$$

として，楽譜上で同じ相対的な変化を持っているコーパス中の d_o の分布を推定する．同様に入力歌唱音声での d を計算し，これらの値を用いて同じ音高変化を歌ったと考えられる候補を挙げる．

$$P_1=P(d_1|d_o),P_2=P(d_2|d_o),P_3=P(d_3|d_o)\cdots\qquad(4)$$

d は正規分布を成すとみなし，全相対音高差に対する尤度を計算する．

$$p(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

μ …コーパスにおける平均情報

σ^2…コーパスにおける分散情報　　(5)

とし，値が最大のものを尤もらしい音高変化とする．

2.3. 音高推定

まず，音高を抽出するために自己相関係数を求め，近似微分を用いた簡単なピーク抽出を試みた．しかし倍音成分が含まれている音声からの抽出を試みた際に，オクターブエラーが起こってしまい，正確に抽出することが困難であった．音高推定法にはノッチフィルタを用いた音高推定[5]や二重くし形フィル タによる独唱・二重唱の音高推定[6]などが挙げられるが，本研究では YIN という手法で基本周波数を抽出することを決めた．

YIN とは自己相関関数を中心に，より正確に基本周波数を推定することを目的としたものである[7]．6 つのステップがあり，ステップ毎の誤り確率は Step1 から Step6 で 10%から 0.5%まで低下している．音高，音高情報の非周期性の割合，その値によって平滑化された瞬間的なパワーの推定結果がそれぞれプロットされ，仕様として，求められた周期点が(6)式によって計算されている．

$$x=\log_2\frac{f}{440}\qquad(6)$$

これを利用して全音高を cent 値で計算し，統計的な処理を行う．cent とは 1 オクターブを 1200 分割したものを表す単位で，相対値を示すのに有効なため(7)式で計算した．

$$\text{cent}=1200*\log\frac{f}{fo}/\log 2\qquad(7)$$

(fo…基準音高, f…対象音高)

男女で同じ楽譜を歌った場合，周波数には 1 オクターブの差があるが，本手法では相対的な値を用いるので考慮しなくて良い．よって基準は C4(261.62Hz)とした．

本研究では一定音高区間における平均音高などをみるため，音符への分割を行う必要がある．本研究での一定音高区間は次のようにする．音高情報のみを用いた場合，楽譜上では違う音だが歌唱者が続けて同じ音で歌唱している部分を検知出来ない．YIN では，音高情報の非周期性の割合と，その値によって平滑化された瞬間的なパワーについても計算がされているため，セント値に加えこの二つの値を用いることで，時間区間の検出を行うこと

C5	B4	A4	G4	A4	G4	A4	B4	C5
1200	1100	900	700	900	700	900	1100	1200

図 3　英語表記例

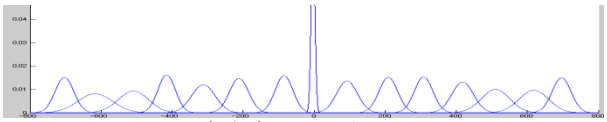

図 4　音高変化の分布(M001,80bpm)

表 1　コーパス使用曲の内容.
(拍…拍子，曲…曲名，発…発音数)

拍	2/4					3/4				
曲	13a	18d	19a	25a	30f	17e	18b	19b	25c	30c
発	21	30	24	45	128	31	66	30	38	65

が出来る．前者は非周期的な(音高が定常的でない)部分，後者は歌唱がある(入力がある)部分で値が大きくなる．

3. 音高に対する統計分析

相対音高変化のずれについての傾向をみるため，まずある程度の楽曲を歌唱した中から楽譜情報上で等しい変化を持つもの同士での分類を行い，その分布をみた．

3.1. バランス歌唱コーパスの構築

相対音高差ごとの分布をみるため，多様な変化を含む楽曲を持つ歌唱コーパスが必要となる．「歌唱」は「話す」行為に比べて制御が困難なため，一被験者に対して多くのサンプルを収録することが難しい．よって推定の際に必要な相対音高差を，少ない曲数で効率よく満たしている必要がある．この条件を満たしていると考えられる歌唱練習曲集の中の「コールユーブンゲン」から抽出することを決めた．

はじめに，収録楽曲の選定について述べる．歌謡曲は 2 拍子，3 拍子，4 拍子が多く見られ，4 拍子は 2 拍子の組合せで表わせるため，楽曲は 2 拍子と 3 拍子とした．また多数楽曲への対応のため，テンポ(bpm)についてもいくつか収録する必要がある．本研究では 10 曲(2 拍子,3 拍子計 5 曲)，テンポは 80bpm,120bpm とし，2 人の収録を行った．なお，データは階名(イタリア音名)での歌唱となっている．収録に際して次のような点を考慮することとした．

・収録の流れ
・収録時間の考慮
・波形データの管理

コーパスの構築に用いた曲の内容を表 1 に示す．これらは全て ト長調の調性で記述されているため，今回の実験では基準音を与え，ト長調で歌唱させた．

3.2. コーパス・サンプルの収録

コーパスに加え，評価用のサンプルの収録も行った．

収録環境

・録音ソフト：YAMAHA SOL2
・波形編集ソフト：YAMAHA TWE
・マイク：SONY F-V320

表 2　相対音高変化のラベリングとグループ分け.
(ラ…ラベル，相…相対差，変…変化数)

ラ	1	2	3	4	5	6	7	8
相	-700	-600	-500	-400	-300	-200	-100	0
変	21	10	31	21	31	83	33	8
ラ	9	10	11	12	13	14	15	
相	100	200	300	400	500	600	700	
変	34	81	32	22	31	9	21	

表 3　正解精度

80bpm	**総数**	**正**	**誤**	**正解精度**
004a	14	14	0	1.000
031b	32	28	4	0.875
031e	31	29	2	0.935

データ形式

・WAVE(16bit,44.1kHz,モノラル)

評価用サンプル

・「コールユーブンゲン」
4-a,13-a,18-b,30-f,31-b,31-e
・テンポ：80bpm,120bpm
・被験者：2 名(M001,M002)

2.3 で述べたものを基に波形データを一音ごとに切り分ける．また子音には基本周波数を持たないものがあるため，母音部分のみを計算に用いることとした．以後はこのデータを用いて実験を行う．

3.3. 曲中における音高変化の数値化

3.2.で得た数値と楽譜情報との比較で分析を行うが，楽譜情報をそのまま処理に用いる事は出来ない．よって楽譜情報の音程を数値化する事を考える．まずコーパスの 10 曲を英語表記で表し，C4 を基準とした cent 値による数値化を行い(図 3)，また隣接する二つの音高の差を計算，表に表し，変化ごとにグループ分けを行った(表 2).

コーパスにおける尤度を計算する際に必要な平均と分散をラベルごとに計算した結果，図 4 を得た．左から，ラベル 1 から 15 までの分布となっている．

4.提案手法の評価

本章では尤度による相対音高差の推定手法について評価する．まず本手法の相対音高変化の推定精度評価のため，一つのコーパスを用いて同一歌唱者のサンプルによる相対音高変化推定の実験を行った．次にコーパスに含まれない歌唱者について同様に推定可能かをみるため，他者の学習データを用いた推定の実験を行った．最後に収録テンポによらず推定可能かの評価のため，異なるテンポの学習データを用いた推定実験を行った．

4.1.入力歌唱における相対音高変化の推定

収録した M001,80bpm のコーパスを用い，M001,80bpm の歌唱音声について実際に尤度を用いた相対音高変化の推定を行った．評価用サンプル No4a,No31b,No31e を使用し，推定結果は相対変化のラベルに基づいたものとした．結果最大で 100%，最低で約 86%の精度を得られた(表 3).

表 4　正解精度(他者コーパス使用)

	004a	**031b**	**031e**
002-001	1.000	0.656	0.839

表５　正解精度(他テンポコーパス使用)

	004a	**031b**	**031e**
80-80	1.000	0.875	0.935
120-80	1.000	0.813	1.000
80-120	1.000	0.844	0.871
120-120	1.000	0.781	0.871

誤推定されたものについては，概ね平均が近いラベルと認識されていた．

4.2.他者のコーパスを用いた推定

他者のコーパスを用いた推定が可能かを見るために，M002,80bpm の歌唱コーパスを用いた M001,80bpm 歌唱の相対音高変化推定を行った．004a は変わらず 100%の精度を得たが，031b，031e に関しては精度が最低で約 66%に落ちている(表４)．よってこのモデルでは他者のコーパスで推定可能なものとそうでないものがあると言える．

4.3.異なるテンポのコーパスを用いた推定

入力歌唱と異なるテンポのコーパスを用いての推定が可能かを見るために，それぞれ

・M001,80bpm の歌唱コーパスを用いた M001,80bpm
・M001,120bpm の歌唱コーパスを用いた M001,80bpm
・M001,80bpm の歌唱コーパスを用いた M001,120bpm
・M001,120bpm の歌唱コーパスを用いた M001,120bpm

の推定を行った．004a は全て推定が可能であり，31e もそれほど変わらない精度を得ているが，31b は 120bpm のコーパスを用いた場合に 80bpm のコーパスを用いた場合よりも精度が約 6%下がっている事がわかる(表５).

5. 考察

5.1. 他者コーパスを用いた推定に関する考察

031b,031e 共に精度が落ちたが，004a については下がらなかった．そこで曲の難度についての考察を行った．

表 6 の平均変化値は，曲中の相対変化値の絶対値の平均を取ったもので，値が大きくなるほどその曲が大きな変化を持っていると言える．この表と楽譜から，曲が長く，大きな変化をもつものほど精度が落ちている事が分かる．よって曲が短く，また大きな変化がない曲であれば，他者のコーパスを用いた推定が出来る可能が大きいが，個人によって歌唱難度が変わるような場合には各々の歌唱コーパスを用いた推定が必要となるといえる．

5.2. 他テンポコーパスを用いた推定に関する考察

80bpm と 120bpm の違いの有無をみるため，歌唱コーパスにおける平均と分散を比較する(図５)．分散が広いほど不安定であり，歌唱が難しい事を示す．よって 80bpm よりも 120bpm の歌唱が難しく，また差がある事が分かる．

また 4.3.に加え 013a,017e,018b,019a の 4 曲に関し同様の

表６　曲の変化と精度差

	004a	031b	031e
平均変化値	171.429	406.250	332.258
総変化数	14	33	32
精度差	0	0.219	0.096

表７　他テンポコーパス使用による正解精度

	013a	**017e**	**018b**	**019a**
80-80	1.000	0.900	0.985	0.870
120-80	1.000	0.867	0.985	0.826
80-120	1.000	0.933	0.877	0.913
120-120	0.950	0.967	0.892	0.870

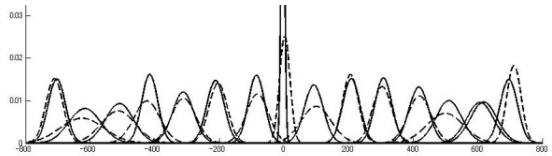

図 ５　80bpm,120bpm の分布(**青**:80bpm,**赤**:120bpm)

追加実験を行った(表７)．80bpm での推定精度が高い部分があり，それ以外でもそれほど変らないため 80bpm のみのコーパスで 他テンポに対応出来る可能性 がある．

6. むすび

本研究では尤度を用いた相対音高変化の推定法を提案した．同歌唱者,同テンポコーパスを用いた推定では，正解精度として最大 100%が得られ，また 80bpm コーパスでの 120bpm 推定は，80%以上の精度が得られた．検定や推定結果から，一つのコーパスデータから他者の歌唱についても推定できるものもあるが，曲ごとに歌唱の難度が異なり半音変化などに差が表れているため，個々のコーパスを用意して推定を行うべきものもあるという事が分かった．今後は，休符を含めた前後の関係や，多様なテンポのサンプルを用いた評価，歌唱モデルの具体的階名推定への拡張について検討し，テンポの種類に限らずさらに多くのサンプルに対しての評価を行いたい．

文　献

[1] 蔭山ほか, “ハミング歌唱を手掛りとするメロディ検索,”信学論，Vol.J77-D-2, No.8(1994) pp. 1543-1551

[2] 清水ほか，“ハミングからの階名と音価の推定 ,” FIT2004 一般講演論文集（同志社大学）,G-016 (2004)

[3] 板橋, “音声工学,” pp. 12-18, 森北出版株式会社,2006

[4] 鈴木ほか, “明解・音楽用語辞典,” p56”平均律”, 株式会社ドレミ楽譜出版社

[5] 中野ほか, “楽譜情報を用いない歌唱力自動評価法,” 情報処理学会論文誌，Vol.48, No.1(2007) pp. 227-235

[6] 干場ほか, “ノッチフィルタを用いた音高推定,” 信学論技報,応用音響, Vol.107, No.62(2007) pp. 31-36

[7] Alain de Cheveigne et.al., “YIN, a fundamental frequency estimator for speech and music,” J. Acoust. Soc. Am., Vol. 111, No. 4, April 2002 pp. 1917-1930.

[8] 山口, “二重くし形フィルタによる独唱・二重唱の音高推定,” 電学論 C, Vol.121-C, No.12 (2001) pp. 1853-1859