Vol.2010-MUS-85 No.10
2010/5/28

歌唱の「熱唱度」評価の検討

大道　竜之介[†1]　伊藤　仁[†1,*1]
伊藤　彰則[†1]　牧野　正三[†1,*2]

歌唱音声の新たな評価指標として「熱唱度」の導入を提案する．聴取実験の結果から，歌唱音声中のビブラートおよび呼吸音が，熱唱の知覚に関与することがわかった．本稿では，歌唱音声におけるビブラート，有声呼吸音，声門破裂の 3 つの特徴を挙げ，これらを定量化する音響特徴量について検討する．34 名の歌唱音声に対する聴取実験から得た，熱唱度の聴取実験スコアと，複数の音響特徴量との重回帰分析を行い，それらの間に重相関係数 0.45 を得た．

Measuring "enthusiasm" of singing voice

Ryunosuke Daido,[†1] Masashi Ito,[†1] Akinori Ito[†1]
and Shozo Makino[†1]

We propose introducing "enthusiasm" as a novel index of singing voice. The result of the listening experiment by human subjects suggests that both vibrato and breath sounds in singing voice concern human perception of enthusiasm. This paper describes our experiments to quantify 3 features in singing voice; vibrato, voiced breath sounds and glottal plosion. As a result of the multiple linear regression analysis between perceived enthusiasm score evaluated by the listening experiment with singing voice recordings of 34 people and some quantified acoustic features, we reached multiple correlation coefficient of 0.45.

†1 東北大学
Tohoku University
∗1 現在，東北工業大学
Presently with Tohoku Institute of Technology
∗2 現在，東北文化学園大学
Presently with Tohoku Bunka Gakuen University

1. はじめに

「歌を歌うこと」は，人種，言語，地域，貧富や老若男女を問わない人類共通の文化であり，本来的に人間に備わった機能から生まれる無償のエンタテイメントである．高価な楽器や特別な演奏技術を必ずしも必要とせず，身体そのものを楽器として音楽を奏でる演奏形態によって，歌唱者は意思や感情をじかに音へと昇華させることができる．この点において歌を歌うことは，誰しもが享受できる表現の文化ありながら，あらゆる音楽行為のなかでも抜群の表現力をもつ演奏行為といえる．

歌唱の良さ，すなわち歌のうまさを評価することは古来行われており，生体的な挙動と歌声の関係を科学的に調べる研究も行われてきている[1)]．近年ではコンピュータによる歌唱の良さの自動評価手法も数多く提案されている．歌唱自動評価の重要な応用はカラオケの採点機能であり[2)]，その評価は音程の正しさやビブラートなどの歌唱法を基準として行われている[3),4)]．歌唱の良さは歌の評価の一つではあるが，それはあくまでも歌から聴取者が受ける印象の評価の一つである．そのほかに歌を評価する基準としては「感情」があり[5)]，その音響的な性質を調べたり[6)]，また歌唱音声合成に応用する[7)] などの研究がある．

これらとは異なる歌唱の評価軸として，歌唱者が歌をどれだけ一生懸命歌っているか，すなわちどれだけ「熱唱」しているかという評価軸が考えられる．これは「感情」と似ているが異なるものである．例えば，6) ではプロの歌唱者が「喜び」「怒り」「怖れ」「寂しさ」などの感情を込めて歌った歌唱が対象であり，「どれだけがんばって歌ったか」という評価とは基準が異なる．このような点に注目して歌唱を評価する試みはこれまで例がない．本研究では，この「熱唱」が数値化できるという仮定を置いて，歌唱がどれだけ熱唱なのか，すなわち「熱唱度」を評価することを試みる．

熱唱度の観点として，本研究では「本人熱唱度」と「知覚熱唱度」の 2 つを導入し，これらを明確に区別して扱う．「本人熱唱度」は歌唱者本人の意思に基づく熱唱度，「知覚熱唱度」は歌唱者以外の人が歌唱音声を聞いた際に知覚する熱唱度とする．歌唱音声に対する熱唱度の評価値には，主に知覚熱唱度の評価値を用いる．目指すシステムとして，人が聞いて知覚する熱唱度に近い値を返すシステムが自然であると考えること，また実験でのやや特殊な環境において歌唱音声を収録した場合，指定した熱唱度の値と実際の本人熱唱度とが，必ずしも一致しないと考えられることがその理由である．

本研究の概要は次の通りである．まず，「熱唱」および「普通の歌唱」をするよう歌唱者に指示することによって，さまざまな歌唱音声を収録する．これにより，「本人熱唱度」の

異なる音声が収録できる．次に，必ずしも本人熱唱度と知覚熱唱度が一致しないことから，収録したデータベースの各歌唱音声について，主観評価によって「知覚熱唱度」を求める．さらに，知覚熱唱度の高い音声と低い音声を分析し，熱唱度と関連しそうな現象の分析とその定量化を行う．最後に，有効だと考えられる物理量を使い，重回帰分析によって知覚熱唱度の推定を試みる．

2. 歌唱収録と聴取実験

2.1 歌唱収録

表 1 に示す 34 人の男女に, 表 2 に示す同一のポピュラーソング（『いとしのエリー』）を伴奏付きで歌唱してもらい, 音声のみを収録した．熱唱度のダイナミックレンジをとるため, 各歌唱者には「熱唱」と「普通」の 2 通りの歌唱を各 1 回歌唱してもらった．

2.2 聴取実験

収録した歌唱音声に対して, 人間が知覚する熱唱度を順位付けることを目的として, 下記のように聴取実験を行った．この聴取実験には作為的な操作を加えた刺激を混入し, ビブラートおよび呼吸音が, 熱唱の知覚に関与するか否かを確かめた．

2.2.1 刺激

聴取実験の刺激として, 2.1 節で収録した歌唱音声より, 歌詞「Elly, my love, so sweet.」の「Elly」区間およびその前後のブレス区間（2〜3[sec] 程度）を切り出して用いた（図 1）．この「Elly」区間は収録音声 1 つあたり 4 回現れる.4 回すべて同一のリズム・音程である．今回はこれらの区間をすべて混ぜて用いた．歌唱者 34 名の「熱唱」および「普通」の歌唱より各 4 つの「Elly」区間を切り出し, 計 272 個の刺激を用意した．

さらにビブラートおよび呼吸音の, 熱唱の知覚への関与を調べる目的で, 下記のような作為的な刺激を用意した．なお計 10 個の作為的な刺激に, 歌唱者の重複はない．

- ビブラートを直線化したもの（5 個）
  - 上記 272 個の刺激のうちビブラートが顕著にみられる刺激を 5 個選び, 音声分析・合成アプリケーション Praat[8] を用いて, 基本周波数（F_0）抽出（10[msec] ごと）を行った．ここでビブラート区間の F_0 に対して対数周波数軸上の回帰直線を計算し, Praat を用いて音声を再合成することで, ビブラートを直線化した（図 2）．このときビブラート区間は手動で指定した．
- 呼吸音を無音化したもの（5 個）
  - 上記 272 個の刺激のうち呼吸音が顕著にみられる刺激を 5 個選び, 波形編集アプリケーションを用いて, 呼吸音区間を無音化した．このとき呼吸音区間は手動で指定した．

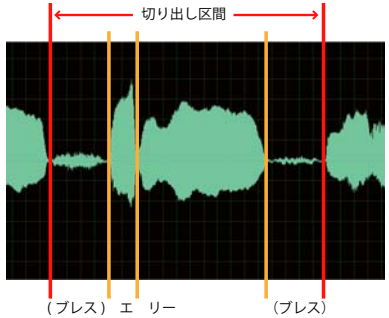


図 1 実験に用いた歌唱音声刺激の切り出し区間

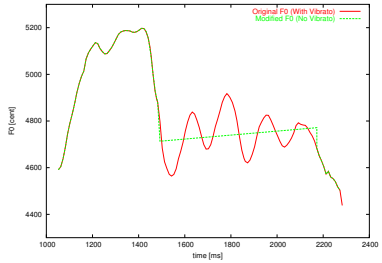


図 2 ビブラートの直線化

この実験においては, 録音レベルやマイクとの距離など収録環境による振幅の差異を吸収するため, すべての刺激の振幅に対する二乗平均平方根（RMS）が一致するよう, 振幅を正規化した．

以上を表 3,4 にまとめる．

2.2.2 被験者および評価方法

大学生の男女 30 人を被験者として, 以下の手順で評価を行った．(2) の評価練習は, 被験者に刺激のダイナミックレンジを提示することを意図して行った．

(1) 各被験者ごとに,282 個の刺激から 100 個の評価用刺激をほぼ無作為に選出する（ただし被験者 30 人の評価によって, 各刺激が評価される回数がほぼ同等になるようにした）

(2) (1) で選出した評価用刺激とは重複のない刺激を 20 個提示し, 評価練習を行う

(3) (1) で選出した評価用刺激を無作為な順で次々に提示して評価を行う

(4) (3) の操作を 3 セット繰り返す（すなわち各被験者は 1 つの刺激を 3 回評価する）

評価語およびスコアとの対応は表 5 の通りとし, 評価用インターフェースでは評価語のみを表 5 と同様に配置して評価を選択してもらった．

各刺激に対して全被験者が付与した評価値の平均を, その刺激の聴取実験スコアとして求めた．

Vol.2010-MUS-85 No.10
2010/5/28

表 1 歌唱音声収録人数

	男性 [人]	女性 [人]	計 [人]
人数	24	10	34

表 2 収録曲

収録曲	『いとしのエリー』1 番のみ
キー/音域	C-Maj. ／ E3〜G4
歌唱部時間	約 1 分 32 秒

表 3 聴取実験の刺激

時間	約 2〜3[sec]
歌詞	「エリー」
音高	男性：C4〜E4 女性：C5〜E5[⋆1]
歌唱者数	34 人 （男性 24 人, 女性 10 人）

表 4 聴取実験の刺激個数

通常の刺激	272 個
ビブラートを直線化した刺激	5 個
呼吸音を無音化した刺激	5 個
計	282 個

表 5 聴取実験の評価語

評価語	スコア
熱唱していると感じる	2
どちらともいえない	1
熱唱していると感じない	0

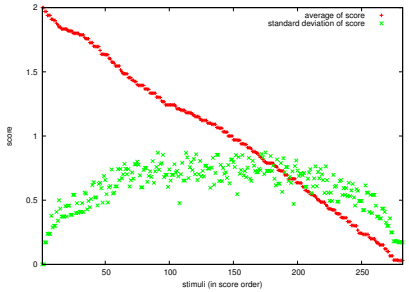


図 3 聴取実験スコア

2.2.3 評 価 結 果

各刺激に対する聴取実験スコアおよび全被験者が付与した評価値の標準偏差を, スコア順位に沿って並べたものが, 図 3 である. 提示した全刺激がスコアに偏りなく一様に評価されていることが確認できる. また知覚熱唱度が極めて大きい, または極めて小さいと評価される音声では安定した評価を得ており, 知覚熱唱度が中程度に評価される音声では評価が不安定となることが見て取れる. なお, ひとりの被験者が与えた評価値の平均は 0.79〜1.55, 評価値の分散は 0.08〜0.32 であった. 被験者から得た内省報告に基づき, 熱唱度を高く感じるとされた特徴と, 熱唱度を低く感じるとされた特徴を表 6 に示す.

⋆1 男性のキーで歌唱した女性もいる

表 6 被験者からの内省報告に基づく特徴

熱唱度を高く感じる	・ビブラートが存在する ・呼吸音がよく聞こえる ・語頭の音程をずり上げる ・語頭の音程ずり上げとビブラートが両方存在する ・やや音程をはずしている
熱唱度を低く感じる	・明らかに音程をはずしている ・声の音色が暗い

3. 特徴の分析

ここでは熱唱の知覚に関与すると推察される歌唱音声の特徴として, ビブラート, 有声呼吸音, 声門破裂を挙げ, 音響的特徴量によってそれらを定量化する手法について検討する.

3.1 ビブラート

表 6 に示した被験者からの内省報告において, ビブラートの存在が熱唱度を高く感じる特徴として挙げられている. ビブラートに関しては多数の研究報告がなされており, 熟達した歌唱者が頻繁に用いるテクニックのひとつであるとする知見が少なくない. また中野らの報告[4] においては歌唱力の自動評価に用いる指標のひとつとして有力であるという. そこでビブラートは歌唱音声を特徴づける指標のひとつとして有力と考え, 音響特徴量の抽出を行った.

ビブラート区間の検出には中野らの方法[4] を用いた. $F_0(t)$[cent] の一次差分 $\Delta F_0(t)$[cent] に対する短時間フーリエ変換（STFT）によって得られる振幅スペクトル $X(f,t)$ に, ビブラートの速さに対応した鋭いピークが現れることを利用する. 表 7 の条件のもと, ある時刻におけるビブラート速さ帯域のパワー $\Psi_v(t)$ と鋭さ $S_v(t)$, さらに「ビブラートらしさ」$P_v(t)$ を次式の通り定義する.

$$\Psi_v(t) = \int_{F_L}^{F_H} \hat{X}(f,t)df \tag{1}$$

$$S_v(t) = \int_{F_L}^{F_H} \left| \frac{\partial \hat{X}(f,t)}{\partial f} \right| df \tag{2}$$

$$sP_v = S_v(t)\Psi_v(t) \tag{3}$$

ただし $\hat{X}(f,t)$ は次式に示す通り, 各時刻での $X(f,t)$ を全周波数帯域のパワーで正規化したもである.

Vol.2010-MUS-85 No.10
2010/5/28

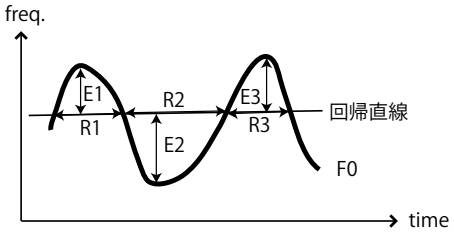


図 4 ビブラート速さと深さの抽出

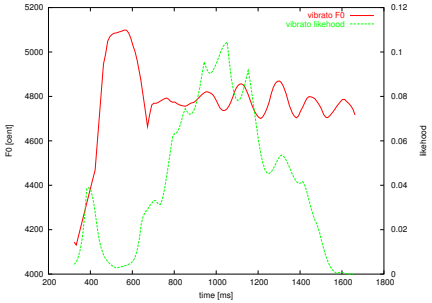


図 5 ビブラートらしさ $P_v(t)$

$$\hat{X}(f,t) = \frac{X(f,t)}{\int X(f,t)df} \tag{4}$$

ビブラートの速さと深さには制限値を設け, それぞれ 5〜8[Hz] および 30〜150[cent] する. 上式での F_H, F_L はビブラート速さの制限値に対応する.

ビブラートらしさ $P_v(t)$ が一定値以上で, その区間の F_0[cent] が, 区間の F_0 回帰直線と 5 回以上交差し[⋆1], なおかつ速さと深さが制限値内である区間をビブラートとして検出したビブラートらしさ $P_v(t)$ の値の例を図 5 に示す.

速さと深さは図 4 および次式のように調べた[⋆2].

$$\frac{1}{\text{rate}} = \frac{1}{N} \cdot \sum_{n=1}^{N} 2R_n \tag{5}$$

$$\text{extent} = \frac{1}{N} \cdot \sum_{n=1}^{N} E_n \tag{6}$$

ビブラートに対する特徴量として, ビブラート検出時間 1, ビブラートらしさの平均 2, ビブラートらしさの最大値 3, 速さ 4, 深さ 5 を抽出した. ただし 2, 3 はビブラート検出区間内で計算した.

3.2 有声呼吸音

表 6 に示した被験者からの内省報告において, 呼吸音は熱唱度を高く感じる特徴として挙げられている. 呼吸音から抽出する音響的特徴量を検討するため, 2.1 節で収録した歌唱音声に含まれる呼吸音を詳細に観察した. その結果, 一部の呼吸音には,200-400[Hz] 付近に, 基本周波数（F_0）のようなピークが見て取れた（図 7）. 高次の調波成分はもたない場合が多く（F_0 のみ）, 存在する場合も $4 \times F_0$ までしか見られない. このピークは呼吸音の後半から末尾にかけて現れる場合が多いようである. このようなピークを持つ呼吸音を有声呼吸音とよぶことにする.

図 7 のようなピークの様相から, 有声呼吸音は, 勢いよく吸気した際に声帯が振動して生じていると考えられる. 平常時の人の呼吸音とは明らかに異なる特徴であり, 一生懸命な歌唱を印象づけると推察し, 有声呼吸音に注目して音響特徴量の抽出を行った.

呼吸音の有声判定には零交差数を用いた. 図 7 に見られる通り, 有声区間においては基本周波数成分が支配的になるため零交差数が小さくなる. 零交差数が一定値以下で, かつ絶対パワーが一定値以上（呼吸音が存在するとみなせる程度）であった区間を有声区間とみなした. 図 6 に図 7 と同一の呼吸音について調べた零交差数の時系列を示す.

計算条件は表 8 の通りとし, 特徴量として有声区間の平均パワー 6 を抽出した.

なおこの計算においては呼吸音の低周波成分を取り除く目的で, まず歌唱区間および呼吸区間を含む音声全体にハイパスフィルタ（カットオフ 80Hz, 遷移幅 20Hz）をかけ, 全音声の RMS 値が一定となるよう正規化した後, ブレス音区間を手動で切り出した. 歌唱音声中の呼吸音区間の検出手法については中野ら[9] が提案しており, 将来的にはこのような方法を用いて自動検出することを目指す.

3.3 声門破裂

2.1 節で収録した歌唱音声を観察したところ, 一部の歌唱において, 母音で始まる言葉を歌う際に声門破裂らしき特徴がみられた. 図 8 はその一例であり, 振幅が急激に立ち上がっている. このように急激に振幅が立ち上がる音声には熱唱度を高く知覚すると推察し, Δ パワーを用いて振幅の立ち上がりの鋭さを評価することを考えた.

Δ パワー（ΔP_i）の計算には Furui[10] による式 (7) を用いた.

表 7 ビブラート計算条件

F_0 間隔	10[ms]
FFT 点数	32[点]
シフト	1[点]
窓関数	ハニング窓
速さ制限値	5〜8[Hz]
深さ制限値	30〜150[cent]

表 8 呼吸音の零交差数計算条件

サンプリング周波数	44.1[kHz]
フレーム幅	10[msec]
フレームシフト	10[msec]
零交差数閾値	4600[回/sec]

⋆1 4) では回帰直線でなく, 平均 F_0 との交差数を用いている. 本実験では 2.2.1 節の作為的な刺激との統一のため回帰直線を用いた.
⋆2 4) での定義とは異なる

表 9　Δ パワー計算条件

サンプリング周波数	44.1[kHz]
フレーム幅	20[msec]
フレームシフト	10[msec]
サイドフレーム（n_0）	4[フレーム]

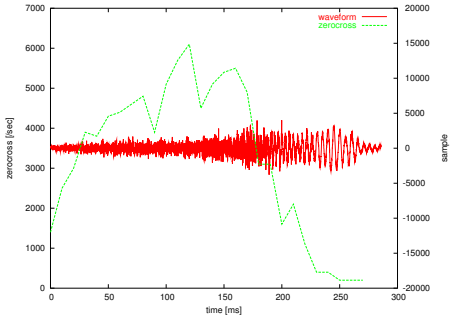


図 6　有声呼吸音（図 7）に対する零交差数

$$\Delta P_i = \frac{\sum_{n=-n_0}^{n_0} P_i(n) \cdot n}{\sum_{n=-n_0}^{n_0} n^2} \tag{7}$$

計算条件は表 9 の通りとし, 特徴量として各音声中の Δ パワーの最大値 7 を抽出した.

4. 評価実験と分析

4.1 聴取実験の結果

作為的な刺激と, その元となった刺激の評価値の平均を図 9(a),(b) に示す. 誤差棒は標準偏差を示す. 呼吸音を無音化したもの, ビブラートを直線化したものともに, 作為的な刺激はすべて, 元となった刺激よりも聴取実験スコアが低い. この結果は, 呼吸音とビブラートの存在が知覚熱唱度を向上させることを示しており, 表 6 に示した被験者からの内省報告に一致する.

4.2 聴取実験スコアと特徴量との相関

聴取実験に用いた刺激のうち, 作為的な刺激を除いた 272 個の音声について, 1 ～ 7 の特徴量を計算した. これらの特徴量と聴取実験スコアとの相関係数, および特徴量間の相関を表 10 に示す.

4.3 重回帰分析

1 ～ 7 の特徴量を説明変数, 聴取実験スコアを目的変数とした重回帰分析を行った. まず 7 個全ての特徴量を説明変数として用いて分析した. ここで各説明変数の, 回帰式に対する影響の強さとして t 値を求め,t 値の絶対値が小さいものから順にひとつずつ省いて重回帰分析を繰り返した. 両側危険率 5%の t 検定において, すべての説明変数が有意と検定された時点で分析を終えた.

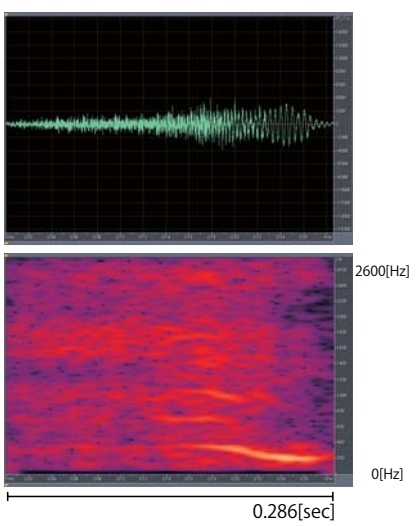


図 7　有声区間を含む呼吸音

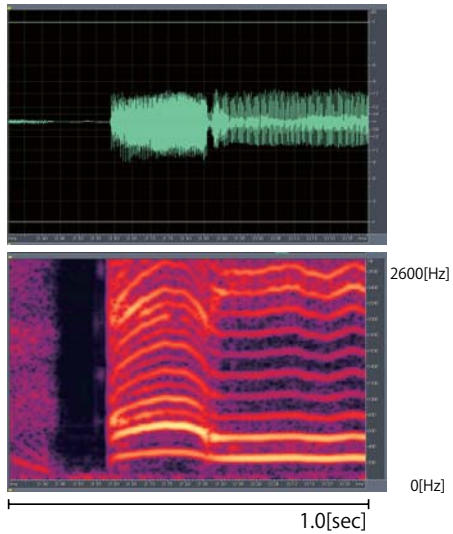
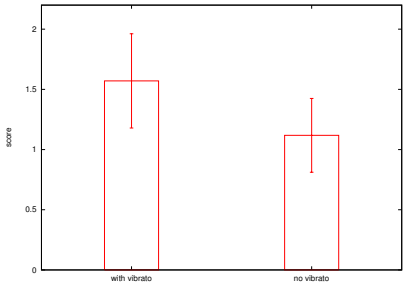
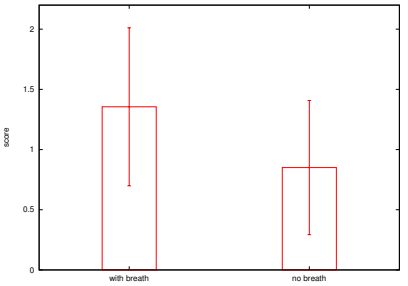


図 8　声門破裂の例



(a) ビブラート直線化の影響



(b) 呼吸音無音化の影響

図 9　作為的な刺激と元の刺激のスコア変化

表 10 特徴量間および聴取実験スコアとの相関係数

			ビブラート					有声呼吸音	声門破裂
		聴取スコア	時間	らしさ平均	らしさ最大	速さ	深さ	有声区間パワー	Δ パワー最大
ビブラート	聴取スコア	1.00							
	時間	0.32	1.00						
	らしさ平均	0.30	0.94	1.00					
	らしさ最大	0.29	0.94	0.97	1.00				
	速さ	0.16	0.71	0.72	0.73	1.00			
	深さ	0.35	0.66	0.69	0.72	0.48	1.00		
有声呼吸音	有声区間パワー	0.08	-0.15	-0.15	-0.15	-0.13	-0.13	1.00	
声門破裂	Δ パワー最大	0.23	0.13	0.11	0.12	0.10	0.10	-0.27	1.00

表 11 重回帰分析の結果

重相関係数	0.45

説明変数	係数	標準誤差	t 値	P 値
切片	$3.51{\cdot}10^{-02}$	$1.78{\cdot}10^{-01}$	$1.98{\cdot}10^{-01}$	$8.44{\cdot}10^{-01}$
ビブラート時間	$2.39{\cdot}10^{-04}$	$1.19{\cdot}10^{-04}$	2.01	$4.50{\cdot}10^{-02}$
ビブラート深さ	$4.01{\cdot}10^{-03}$	$1.16{\cdot}10^{-03}$	3.45	$6.54{\cdot}10^{-04}$
呼吸音有声区間パワー	$5.55{\cdot}10^{-03}$	$1.63{\cdot}10^{-03}$	3.41	$7.55{\cdot}10^{-04}$
Δ パワー最大	$1.28{\cdot}10^{-01}$	$3.03{\cdot}10^{-02}$	4.22	$3.33{\cdot}10^{-05}$

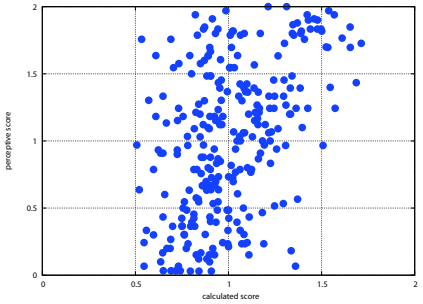


図 10 散布図（客観評価スコア-聴取実験スコア）

この時点で説明変数として残った特徴量は, ビブラートの検出時間 1, ビブラートの深さ 4, 呼吸音の有声区間平均パワー 6, Δパワー最大値 7 の 4 つであった. このときの統計値を表 11 に示す. さらに回帰式から得た客観評価スコアと聴取実験スコアとの関係を, 図 10 の散布図に示す.

4.4 考　　察

いずれの特徴量においても, 聴取実験スコアとの相関係数の値は大きくないが, 複数の特徴量を用いた際の重相関係数は単一の特徴量よりも大きな値が出ており, 相関ありとみなせる範囲に達している. 人間が熱唱度を知覚する特徴は多数あり, そもそも単一の特徴量から得られる相関は大きくないものと考える. 今後適切な特徴量を追加することで, より相関が高い回帰式を得られる見込みがある.

また図 10 から, 客観評価スコアがおよそ 1.0 より大きい範囲では, 客観スコアから聴取スコアの最低値を予測できていることが見て取れる. 各特徴量に対する回帰係数はいずれも正であり, 最終的に用いた 1, 4, 6, 7 の特徴量は, 知覚熱唱度を向上させる特徴指標と認めることができる.

今後, 評価に用いる特徴を増やすこと, また今回注目した特徴とは逆に知覚熱唱度を低下させる特徴について調べ, それらに基づいた特徴量を導入することで, 聴取実験スコアとの相関がさらに高い客観評価を得られると考えらる.

5. む　す　び

本稿では, 歌唱音声の新たな評価指標として「熱唱度」の導入を提案した. 本研究は初期段階にあり, 聴取実験による評価と今回用いた手法による客観評価との相関は, 現時点で高くない. しかし新たな特徴への着目による改善の余地は無数にあり, 将来的には人間の感覚に近い熱唱度評価が可能であると期待する.

参　考　文　献

1) Johan Sundberg:“歌声の科学”, 東京電機大学出版局, 1987.
2) 北村秀仁:“ビブラート採点機能を有するカラオケ採点装置”, 公開特許 2004-102146, 2004.
3) 竹内 英世, 保黒 政大, 梅崎 太造:“カラオケ採点用の高分解能ピッチ抽出法”, 電学論 C , Vol. 129, No. 10, pp.1889-1901, 2009.
4) 中野倫靖, 後藤真孝, 平賀譲:“楽譜情報を用いない歌唱力自動評価手法”, 情報処理学会論文誌 Vol.48 No.1, Jan. 2007.
5) K. R. Scherer:“Expression of emotion in voice and music,”Journal of Voice, vol. 9, no. 3, pp. 235-248, 1995.
6) S. Jansens, G. Bloothooft and G. de Krom:“Perception and acoustics of emotions in singing, Proc Eurospeech,” vol. IV, 2155-2158, 1997.
7) X. Rodet,“Synthesis and processing of the singing voice,” Proc. IEEE Benelux Workshop on Model based Processing and Coding of Audio, 2002.
8) Boersma, Paul:“Praat, a system for doing phonetics by computer. Glot International 5:9/10, 341-345, 2001.
9) 中野倫靖, 後藤真孝, 緒方淳, 平賀譲:“無伴奏歌唱におけるブレスの音響特性とそれに基づく自動ブレス検出”IPSJ SIG Technical Report 2008-MUS-76(15)
10) Sadaoki Furui:“On the role of spectral transition for speech perception,” J. Acoust. Sec. Am., Vol.80, No.4, October 1986