論　　文 43.75.Rs

歌声らしさの知覚モデルに基づいた歌声特有の音響特徴量の分析*

齋藤　毅*1,† 辻　直也*1,†† 鵜木祐史*1 赤木正人*1

[要旨] 歌声特有の音響特徴量と歌声知覚の関係を検討するために，歌声らしさの知覚モデルを提案する。このモデルは，「歌声らしさという聴覚印象が複数の基本的な心理的特徴の知覚に起因する」という仮説のもと，歌声らしさと音響特徴量の対応関係の間に基本的な心理的特徴を介した3層で構成される階層構造モデルである。第1層（歌声らしさ）と第2層（基本的な心理的特徴）の関係については，多次元尺度構成法と重回帰分析によって調査した。第2層と第3層（音響特徴量）の関係については，STRAIGHT を用いた音響分析・合成と心理物理実験によって調査した。その結果，“揺れ,” “響き” といった基本的な心理的特徴が歌声らしさの聴覚印象に大きく寄与しており，両者の聴覚印象には基本周波数の準周期的な振動成分であるヴィブラートとそれに同期したホルマントの振幅変調成分，及び3kHz 付近の顕著なスペクトルピーク成分と同帯域の強い高調波成分がそれぞれ寄与していることが明らかとなった。更に，これらの音響特徴量を話声に付与することで歌声らしさの聴覚印象が向上する結果を得た。以上から，歌声らしさの知覚モデルを構築することで，歌声知覚における歌声特有の音響特徴量の役割について詳細に検討することが可能であることを示した。

キーワード 歌声らしさ，知覚モデル，多次元尺度構成法，ヴィブラート，歌唱ホルマント
Singing-ness, Perceptional model, Multidimensional scaling method, Vibrato, Singing formant

1. はじめに

歌を歌うことは，歌詞としての言語情報だけでなく感情や想いといった非言語情報を表出するコミュニケーション手段である。とりわけ，卓越した歌唱技量を持つ声楽家の “歌声らしい声” を聴くことで，時として話声からは得られない大きな感動を覚えることがある。では，歌声らしい声とはどのような声であろうか？ この問題の回答を得るには “歌声らしさ (Singing-ness)” という非言語情報の知覚がどのように行われいるのか検討する必要がある。

歌声らしさという聴覚印象には，話声にはない歌声特有の音響特徴量が強く寄与していると考えられる。そのため，心理量である歌声らしさと，物理量である音響特徴量の関係が明らかになれば，歌声らしい声を定義することができる。また，歌声らしい声の音響構造が明らかになれば，歌声特有の知覚・生成機構の一端が明らかになり，更には歌声合成といった音声アプリケーションの発展にも大きく貢献できると考えられる。

歌声特有の音響特徴量は，多くの先行研究において報告されている。例えば，基本周波数 (以後 F_0 と呼ぶ) の準周期的変動成分であるヴィブラート (vibrato) [1–3] のような F_0 の変動成分 [4–6] や，3kHz 付近の顕著なスペクトルピーク成分 [7, 8] である歌唱ホルマント (singing formant) が代表的である。これら特徴の一部は，歌声の声質に影響を与える音響特徴量であると報告されており [9]，かつオペラを中心とした洋楽歌唱はもちろん，邦楽歌唱においても存在することが報告されている [10, 11]。

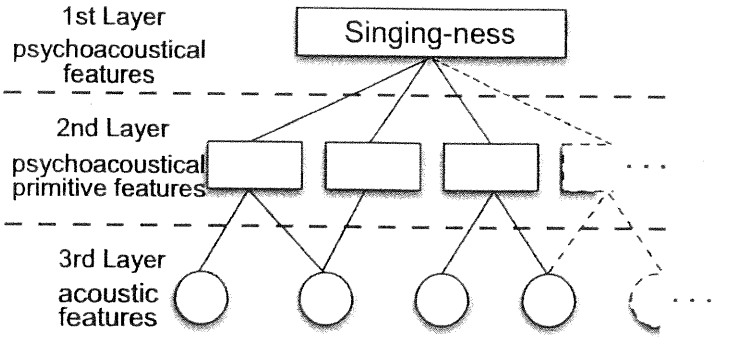
一方で，歌声知覚における心理的特徴に関しても検討されている。西内らは，歌声を評価する際に用いられる表現語に着目し，声の響きや明瞭さといった五つの表現語を歌声知覚において重要な心理的特徴として報告している [12]。また，歌声以外を対象とした取り組みでは，上田による様々な音を対象とした音色の表現語に関する検討 [13] や，木戸らによる通常発話の声質に関連する表現語の抽出 [14] などが行われている。

近年になって，STRAIGHT [15] に代表される高品質な音声分析合成系を用いることで，歌声知覚と音響特徴量の関係が検討されてきている。筆者らは，歌声の F_0 特有の変動成分に着目し，それら成分が歌声の自然性知覚に影響を与えることを示している [16]。また，峯松らによる長唄知覚に重要な音響特徴の検討や [17],

* Analysis of proper acoustic features to singing voice based on a perceptual model of “singing-ness,” by Takeshi Saitou, Naoya Tsuji, Masashi Unoki and Masato Akagi.

*1 北陸先端科学技術大学院大学・情報科学研究科
† 現在，産業技術総合研究所
†† 現在，㈱日本電気
(問合先：齋藤　毅　e-mail: saitou-t@aist.go.jp)

(2007年5月7日受付，2007年11月7日採録決定)



図–1 歌声らしさの知覚モデルの概念図

河原らによるモーフィング合成を用いた歌声の個人性知覚と音響特徴量の関係の検討 [18] など多様な研究が展開されてきている。

STRAIGHT の出現は，高精度な音響分析，自由な音響パラメータ操作，そして高品質な合成音声の生成といった一連の処理を可能にし，歌声知覚研究の発展を加速させている。しかし，歌声らしさという聴覚印象に寄与する音響特徴量に関しては，検討されていないのが現状である。これは，高次の心理的特徴である歌声らしさと物理量である音響特徴量の対応関係を直接的かつ定量的に調査することが，高品質な音声分析合成系をもってしても困難だからである。そのため，歌声らしさの知覚と音響特徴量の関係を新たな枠組みで捉えた分析を行う必要がある。

そこで本論文では，図–1 に示した歌声らしさという心理量と物理量である音響特徴量の関係を階層構造で記述した歌声らしさの知覚モデルを提案し，このモデルに基づいた歌声特有の音響特徴量の分析・解明を行う。このモデルは，歌声らしさの聴覚印象が複数の基本的な心理的特徴の知覚によって構成されるという考えに従い，歌声らしさ（第 1 層）と音響特徴量（第 3 層）の関係を基本的な心理的特徴（第 2 層）を介した 3 層から構成される。隣接する層間の関係を調査することで，これまで直接的な分析が困難であった歌声らしさという高次の心理量と物理量である音響特徴量の対応関係を段階的に解明することが可能となる。

2. 歌声らしさの知覚モデル

我々は，歌声らしい声を説明する際に，F_0 変化中に 5 Hz の速さで振動するヴィブラートが存在する声，或いは 3 kHz 付近に強いスペクトルピークが存在する声とは言わず，明瞭な声とか響いている声と説明することがほとんどである。つまりは，歌声らしさという聴覚印象は，具体的な物理量ではなく何等かの言葉によって表現される。このため，歌声らしい声とそれを特徴付ける音響特徴量の関係は，両者の間に言葉を介して議論することが自然と考えられる。

以上の考えを基に，本論文では歌声らしさの聴覚印象と音響特徴量の関係を記述する歌声らしさの知覚モデルを提案する。図–1 にモデルの概念図を示す。このモデルは 3 層から成る階層構造を持ち，第 1 層は高次の心理的特徴である歌声らしさ，第 2 層は歌声らしさという心理的特徴を説明する言葉（基本的な心理的特徴），そして第 3 層は基本的な心理的特徴の知覚に寄与する音響特徴量からそれぞれ構成される。

本論文では，歌声らしさの知覚モデルを構築することで歌声特有の音響特徴量の分析し，更にはそれら特徴と歌声知覚の関係について詳細に検討する。歌声らしさの知覚モデルは，隣接する各層の関係をトップダウン的に調査することで構築する。第 1 層と 2 層の関係は，多次元尺度構成法による歌声らしさに関する心理空間の構築と，構築した心理空間における基本的な心理的特徴の重回帰分析によって調査する。第 2 層と 3 層の関係は，基本的な心理的特徴の聴覚印象と相関の強い音響特徴量を STRAIGHT を用いた音響分析によって調査する。最後に，抽出した各層の構成要素の妥当性をボトムアップ的に検証することでモデルの評価を行う。以上の取り組みによって，歌声らしさという高次の心理的特徴と音響特徴量の関係を段階的に調査することが可能となり，最終的に抽出された第 3 層の構成要素を歌声らしさの聴覚印象に寄与する音響特徴量として定義できる。

3. 音声データ

本論文で扱う音声データについて述べる。歌声らしさの聴覚印象に寄与する音響特徴量を検討するためには，以下の条件を満たす音声データが必要と考えられる。

- 特定の歌唱法ではなく，様々な歌声を対象にした歌声らしさの知覚について検討するため，多様な歌唱法の歌声であること。
- 話声にない歌声特有の音響特徴量を明確にするため，同一発声者によって発声された話声と歌声が含まれ

ること。

これには，様々な歌唱法，歌唱者による歌声及び話声を対象に歌声らしさの聴覚印象度を調べ，印象度の大きさを基準に，本実験で扱う音声データを選定する必要がある。そこで，本章では，歌声らしさの聴覚印象度を調査するための聴取実験，及びその結果を踏まえた音声データの選定を行う。

3.1 音声データベース

大規模な歌声データベース「日本語を歌，唄，謡う」[19]に収録された様々な歌唱法による歌声及び話声を用いた。歌唱法は，洋楽（ソプラノ，メゾソプラノ，アルト，テノール，バリトン，バス），わらべ歌，民謡，長唄，小唄，琵琶楽，歌舞伎，能，狂言，地歌，清元節，一中節，琉球古典音楽，詩吟，声明，新劇朗読，落語，そしてアナウンサの計18種類（パートの違いも考慮すると計23種類）で，発声者は38人である。データベースには，共通歌詞“かえでいろづくやまのあさは”を朗読及び歌唱した音声が収録されているが，発声者や歌唱法によってメロディが異なることから，日本語5母音の孤立発話（アナウンサ以外の発声者は，歌唱発声と通常発声の音声が収録されている）中の母音/a/を音声データとして用いた。ここで，対象音声を母音/a/に限定したのは，歌声らしさの聴覚印象評定や後に行う音響分析における条件を統一するためである。なお，切り出した母音/a/の発話時間と基本周波数はデータ毎に異なり，発話時間長は最も長いもので3,249 ms，短いもので398 msで，平均基本周波数は最も高いもので542 Hz，低いもので136 Hzであった。

3.2 聴取実験

上記の80種のデータに関する歌声らしさの聴覚印象度を聴取実験によって調査し，より歌声らしいものからより話声らしいものへと順位付けを行った。

実験は，絶対評価法[20]によって行い，その際の評価尺度は，5段階評価尺度（+2：非常に歌声らしい，+1：歌声らしい，0：どちらとも言えない，−1：話声らしい，−2：非常に話声らしい）を用いた。聴取者に評価尺度に慣れてもらうための予備実験を行った後，80種の音声データそれぞれに対する評価実験を3回ずつ行った。なお，いずれの音声データも振幅レベル等の調整は一切行っていない。

聴取者は，正常な聴力を有した大学院生11名（男性10名，女性1名）である。各聴取者の音楽経験には差があるが，いずれも歌唱訓練等を受けた経験はない。

実験は防音室内で行い，聴取者にはヘッドホンを介して音声データを呈示した。その際の音圧レベルは，聴取者の聞き易い大きさに設定した。以下に，使用した実験機器を示す。

音声刺激呈示用サーバ：DAT+LINK & LinuxPC
D/A 変換器：STAX DAC-TALENT BD
ヘッドホン：STAX SR-404
ヘッドホンアンプ：STAX SRM-1 MK-2

3.3 実験データの選定

聴取者全員の各音声データに対する3回の評価平均値から歌声らしさの順位を決定し，その順位を基に本実験で扱う音声データを選定した。**表–1** に選定した11種の音声データを示す。11種の音声データの内訳は，順位が上位のもの（表中のNo. 1〜No. 3：ほとんどの聴取者が歌声らしいと判断）から3種，下位のもの（No. 9〜No. 11：話声らしいと判断）から3種，そして中間の順位だったものから5種（No. 4〜No. 8）である。

音声データの選定においては，洋・邦楽の多様な歌

表–1 本実験で使用する音声データ（○，△，□の記号はそれぞれが同一発声者による音声であることを示す）

音声データ番号（歌声らしさの順位）	種別	歌唱法	性別	平均基本周波数 [Hz]	発話時間長 [ms]
No. 01 ○	歌声	テノール	男性	273	2,093
No. 02	歌声	バリトン	男性	251	1,782
No. 03	歌声	メゾソプラノ	女性	553	1,201
No. 04	歌声	民謡	男性	290	2,599
No. 05	歌声	わらべ歌	女性	547	1,728
No. 06 △	歌声	ソプラノ	女性	382	1,340
No. 07 □	歌声	長唄	男性	209	1,745
No. 08	歌声	声明	男性	175	2,257
No. 09 ○	話声	—	男性	171	517
No. 10 △	話声	—	女性	256	951
No. 11 □	話声	—	男性	195	436

唱法を含み，かつ同一発声者の歌声と話声を含むように考慮した。これにより，歌唱法の種類に依存せず，かつ話声にはない歌声固有の音響特徴の抽出が可能になると考えた。また，アナウンサの音声に関しては，歌声らしさの順位は 80 データ中で 2 番目に低かったものの，その音響特性に関しては通常の話声とは異なることが報告されている [21] ことから，対象音声から除外した。以後，選定した 11 種の音声データを音声刺激と呼ぶ。

歌声らしさの聴覚印象を評定する際に，声の高さや長さの影響が大きいことが考えられるため，表–1 における 11 種の音声刺激を平均基本周波数の高さ，及び発話時間長でそれぞれ順位付けしたものと歌声らしさの順位の相関を調査した。その結果，平均基本周波数で 0.564，発話時間長で 0.603 とさほど高い相関値を示さず，歌声らしさの聴覚印象評定が基本周波数の高さ及び発話時間長以外の様々な音響特徴量の知覚に基づいて行われていることが確認された。そこで，本論文では，平均基本周波数や発話時間長といった大局的な音響特徴量以外で，歌声らしさの聴覚印象に寄与する特徴量について検討する。

4. 歌声らしさの知覚モデルの構築—第 1・2 層間の関係—

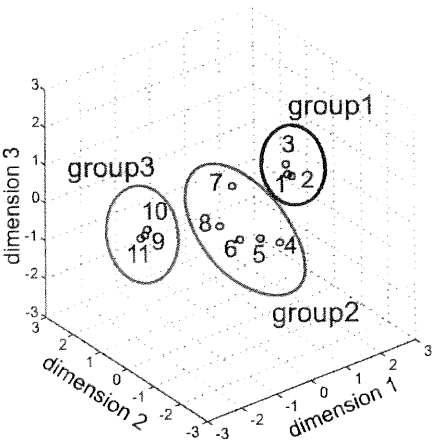
歌声らしさの聴覚印象と，それに寄与する心理的特徴の関係を検討する。最初に，歌声らしさの聴覚印象における音声刺激の関係を多次元空間に布置する。次に，その空間において歌声らしさの聴覚印象に寄与する心理的特徴について検討する。

4.1 歌声らしさの多次元空間

ある心理的特徴を 1 次元でなく多次元で扱う方法の一つに，多次元尺度構成法（MDS: Multidimensional Scaling）がある。西内らは，MDS を用いることで歌声の音色を表す表現語を抽出し，その有効性を示している [12]。そこで，聴取実験によって音声刺激間の歌声らしさに関する類似度を求め，MDS によって各音声刺激の心理空間における布置を決定する。本論文では，構築された心理空間を歌声らしさの空間と呼ぶ。

4.1.1 聴取実験

MDS によって歌声らしさの空間を構築するために，一対比較実験 [20] によって歌声らしさに関する音声刺激の間隔尺度を求めた。実験では，二つの音声刺激を対にして呈示し，どちらがより歌声らしいかを評定させた。刺激対は 11 種の音声刺激から構成され，その数は刺激順序の違いも考慮した 110 組である。なお，聴取者には刺激対の聞き直しは許可しなかった。一対比較における評定尺度は，歌声らしさに関する 5 段階（+2：先の刺激の方が非常に歌声らしい，+1：先の刺激の方が歌声らしい，0：どちらとも言えない，−1：後の刺激の方が歌声らしい，−2：後の刺激の方が非常に歌声らしい）である。なお，聴取者及び実験環境は前節の実験と同様で，音声刺激の振幅レベル調整は行っていない。



図–2 歌声らしさの空間における音声刺激の布置

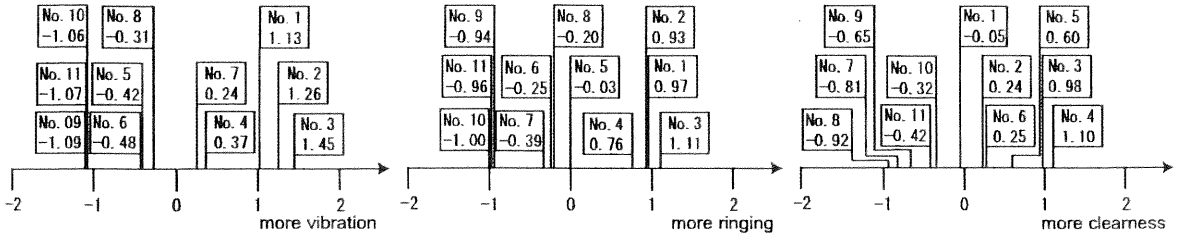
4.1.2 MDS による歌声らしさの空間の構築

聴取実験で得られた歌声らしさに関する音声刺激の間隔尺度を基に，MDS によって歌声らしさの空間を構築した。MDS は，SPSS for Windows による Kruskal の方法 [22] を採用した。その際，次元数と音声刺激の布置の適合度を表す Stress の値は，1 次元で 17.1%，2 次元で 10.7%，そして 3 次元では 5.4%であった。10%以下の Stress 値が適合度として妥当であることから，歌声らしさの空間は少なくとも 3 次元以上で表現することが必要と言える。

図–2 に，3 次元で構築した歌声らしさの空間を示す。この図から，11 種の音声刺激が，グループ 1（音声番号：No. 1，No. 2，No. 3），グループ 2（音声番号：No. 4，No. 5，No. 6，No. 7，No. 8），そしてグループ 3（音声番号：No. 9，No. 10，No. 11）の 3 グループに大別できることが分かる。表–1 の歌声らしさの順位と照合すると，グループ 1 が歌声らしさの順位が高いもの，グループ 3 は順位が低いもの，そしてグループ 2 はその中間の順位の音声刺激から構成されている。この結果から，構築された歌声らしさの空間には，聴取実験で得られた聴取者の聴覚印象が反映されていると考えられる。

4.2 基本的な心理的特徴の導出

歌声らしさの知覚モデルの第 2 層を構成する基本的な心理的特徴として，音色の表現語に着目する。最初に，本研究における歌声らしさの聴覚印象を説明する表現語を選定する。次に，選定した各表現語の聴覚印



図–3 各表現語の聴覚印象に関する音声刺激の間隔尺度（左から揺れ，響き，明瞭さに関する結果）

象に関する音声刺激の間隔尺度を導出する。最後に，重回帰分析によって歌声らしさの空間における各表現語の方向を求めることで歌声らしさの聴覚印象との関係を調査し，歌声らしさの知覚への寄与度が強い表言語を基本的な心理的特徴として定義する。

4.2.1 表現語の選定

3章の歌声らしさの順位付け実験において，聴取者が歌声らしさを評定する際にどのような表現語を用いているのか自由回答形式で記述してもらった結果，揺れ，響き，明瞭さ，という回答が8割以上の聴取者から得られた。また，西内らは歌声を評定する際に用いられる表現語として，響き，明瞭さ，音程の正確さ，音色の統一性，好ましさの計5種類を挙げている [12]。そこで，両実験に共通した表現語である“響き（本論文中では ringing と英訳する）”と“明瞭さ（clearness と英訳する）”に加え，3章の聴取実験ですべての聴取者が回答した“揺れ（vibration と英訳する）”の合計3種を歌声らしさの知覚を構成する基本的な心理的特徴の候補として選定した。

4.2.2 各表現語の聴覚印象の検討

選定した3種の表現語と歌声らしさの関係を調査するために，聴取実験よって各表現語の聴覚印象に関する音声刺激の間隔尺度を求めた。

聴取実験はシェッフェの一対比較法（浦の変法）[20] を採用し，3種の表現語それぞれに関して行った。評定尺度は，各表現語について5段階尺度（例えば揺れの表現語に関しては，+2：先の刺激が非常に揺れている，+1：先の刺激の方が揺れている，0：どちらとも言えない，−1：後の刺激の方が揺れている，−2：後の刺激の方が非常に揺れている）である。呈示した刺激対は11種の音声刺激から構成され，各表現語に関して110組をそれぞれ聴取者に1回のみ呈示した。なお，聴取者及び実験環境は2章の実験と同様で，音声刺激の振幅レベルの調整は行っていない。

各表現語における音声刺激の1次元軸上での間隔尺度を図–3 に示す。図中の番号は，表–1 に示した音声刺激番号に対応し，番号の下に記された数値が大きいほど（水平軸上の右に付置されているものほど），表現語の聴覚印象が強かったことを表す。この結果から，各表現語に関する音声刺激の布置は異なるものの，歌声らしさの順位が高い音声刺激ほど各表現語の聴覚印象が強い傾向が示された。

表–2 歌声らしさの空間における各表現語の方向と重相関係数

表現語		角度 [°]
揺れ–	響き	56.2
	明瞭さ	105.2
響き–	明瞭さ	49.3

表現語	重相関係数
揺れ	0.99
響き	0.99
明瞭さ	0.84

4.2.3 歌声らしさと表現語の関係

各表現語に関する音声刺激の間隔尺度と，歌声らしさの空間における音声刺激の布置の関係を重回帰分析によって調べた。重回帰分析は，図–3 の各表現語における音声刺激の評価値を目的変数，図–2 の歌声らしさの心理的空間における各音声刺激の3次元座標値を説明変数として行った。その際の重回帰式を以下に示す。

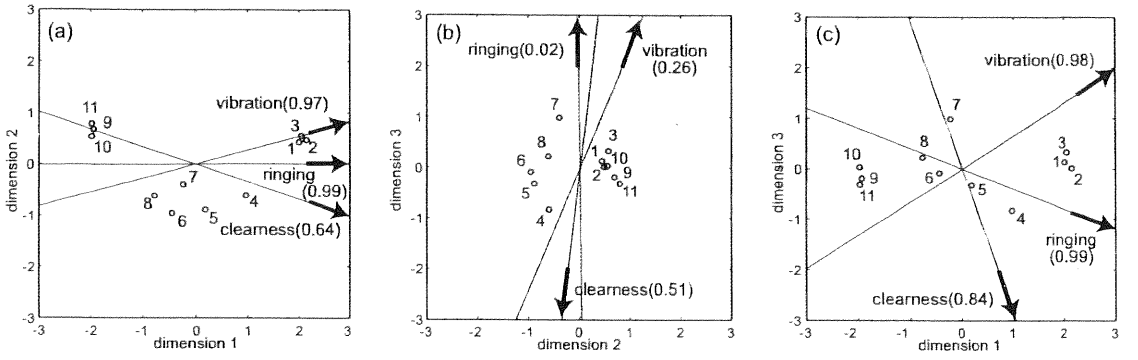
$$y = ax_1 + bx_2 + cx_3 \tag{1}$$

ここで，y は目的変数，x_1，x_2，x_3 は説明変数，そして a，b，c は偏回帰係数である。この式から，

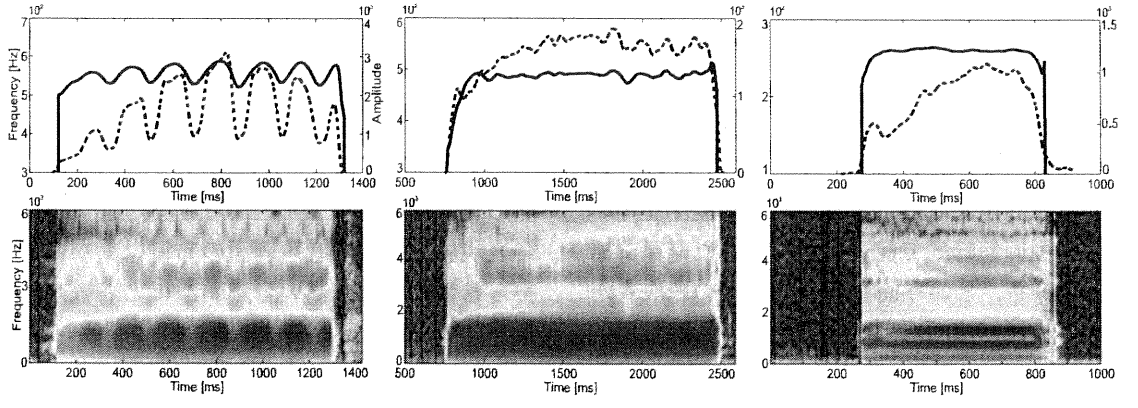
$$E = \sum_{i=1}^{11}(y_i - ax_{1i} - bx_{2i} - cx_{3i}) \tag{2}$$

で表される誤差 E が最小となる偏回帰係数 a, b, c を求めた。ここで，添字 i は，表–1 中の音声刺激番号：1〜11 である。偏回帰係数から得られる歌声らしさの空間中の各表現語の表す方向（角度）と，それぞれの表現語の重相関係数を表–2 に示す。重相関係数の値が 1.00 に近いほど，その表現語が歌声らしさの表現語として適していることを示す。この結果から，すべての表現語が歌声らしさの聴覚印象に寄与しており，とりわけ揺れと響きの重相関係数が大きいことが示された。また，各表現語が示す方向が大きく異なることも明らかとなった。

次に，歌声らしさの空間における音声刺激の布置と各表現語の示す方向を，1–2 平面，2–3 平面，1–3 平面



図–4 歌声らしさの空間における各表現語の方向（左から 1–2 平面，2–3 平面，1–3 平面）



図–5 揺れの聴覚印象度が異なる音声刺激の分析結果
上図の実線は F_0 の時間変化パターン，破線は振幅エンベロープ，下図はサウンドスペクトルグラムを表す（左から揺れの順位が 1 位，7 位，10 位）

のそれぞれに重ね書きしたものを図–4 に示す。図中の各表現語の横に示した数値は，各平面における重相関係数である。すべての平面図において，各表現語の示す方向が大きく異なることが示された。特に，図–4(a), (c) 中では，前節で示した三つのグループが，揺れと響きの指す方向に沿って分布している結果となった。

以上の結果から，揺れ，響き，明瞭さの 3 種が，歌声らしい声を説明する表現語として妥当であることが確認された。中でも，揺れと響きに関しては，歌声らしさの空間における重相関係数が大きく，更には各音声刺激の歌声らしさの聴覚印象が両表現語の示す方向に沿って布置されていることが明らかとなった。これは，2 種の表現語である揺れと響きが歌声と話声を聞き分ける上で重要で，歌声らしさの聴覚印象への寄与度が最も大きい可能性を示唆するものである。以上から，歌声らしさの知覚モデルの第 2 層を構成する心理的特徴として，揺れと響きを採用する。

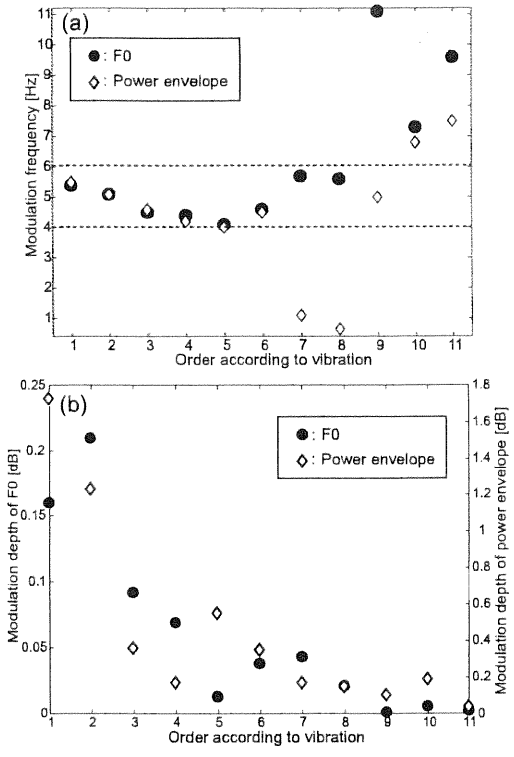
5. 歌声らしさの知覚モデルの構築—第 2・3 層間の関係—

本章では，第 2・3 層の関係の検討として，揺れと響きの聴覚印象に寄与する音響特徴量を抽出する。前章で示した各表現語の聴覚印象に関する評定結果をもとに，評定順位が異なる音声刺激間の音響構造の差異について調査する。

5.1 揺れの知覚に関して

図–3 に示した揺れの聴覚印象に関する評定結果において，評定順位が 1 位（No. 3），7 位（No. 5），9 位（No. 10）の音声刺激を STRAIGHT によって分析した結果を図–5 に示す。ここから，揺れの評定順位が高いものほど，F_0 と振幅エンベロープが大きくかつ周期的に変動し，揺れの評定順位が低くなるに従い変動は小さくなり，その周期性も乏しくなる傾向が確認された。そこで，時間変動が顕著であった F_0 と振幅エンベロープに着目し，その特性について分析した。

図–6(a) は，音声刺激ごとの F_0 と振幅エンベロープに含まれる時間変動の支配的な変調周波数を示したものである。縦軸が支配的な変調周波数，横軸は音声刺激の揺れに関する評定順位を表す。この結果から，評定順位が高い音声刺激において，F_0 と振幅エンベロープがともに 4～6 Hz の変調周波数で振動していることが分かった。逆に，評定順位が低いものは，両者の振動特性が異なり，その特性は 4～6 Hz から逸脱していた。ここで，F_0 における 4～6 Hz の準周期的振



図–6 F_0 と振幅エンベロープに含まれる振動成分（(a)：変調周波数，(b)：偏移幅）

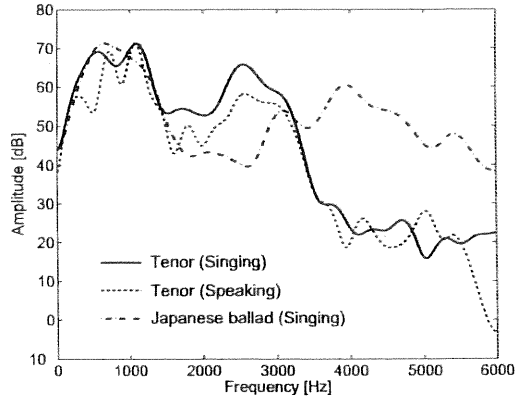
動は，Seashore らが報告 [1] している基本周波数の周期的変動成分であるヴィブラートの特性に近い値である。また，揺れの聴覚印象が 1，2 位の音声刺激における変調周波数は，筆者らの先行研究 [23] において示した歌声合成音に高い自然性を付与するヴィブラート変調周波数：5.5 Hz に近い値である。

図–6(b) は，F_0 と振幅エンベロープの時間変動の 4～6 Hz の周波数帯域における偏移幅を，揺れの順位が高い順に示したものである。この結果から，評定順位が高い音声刺激ほど両者に含まれる変動の偏移幅が大きいことが明らかとなった。

以上から，揺れの聴覚印象に関する評定順位と F_0 及び振幅エンベロープに含まれる準周期的な変動の特性との間には強い相関関係があり，その F_0 変動の特性が歌声特有の音響特徴量であるヴィブラートの特性に似ていることが明らかとなった。また，ヴィブラートに伴った音声振幅エンベロープの変動（AM: Amplitude modulation）も報告されている [3, 17]。以上から，揺れの聴覚印象には，F_0 の準周期変動成分であるヴィブラートと，それに伴う振幅エンベロープの変動という二つの音響特徴量が寄与している可能性が高いことが示された。

5.2 響きの知覚に関して

Vennard は，歌声の「響き」には 2,800 Hz 付近の成分が影響を与えていると報告している [24]。また，

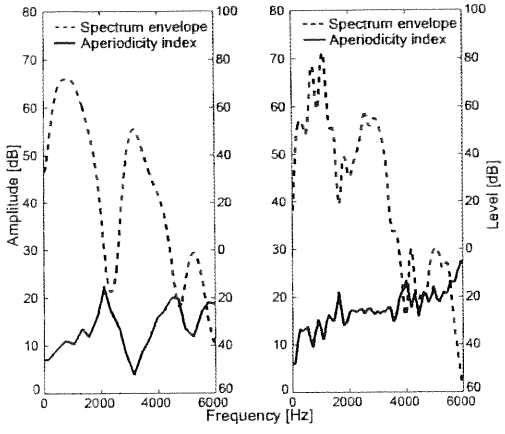


図–7 歌声と話声のスペクトル包絡の比較

Sundberg は，プロの男性オペラ歌手の歌声には，スペクトルの 2～3 kHz 付近において顕著なピークが存在することを示しており [7]，このピークを歌唱ホルマントと呼んでいる。また，オーケストラにおいて歌声が聞こえる要因の一つに歌唱ホルマントの存在を挙げている。邦楽歌唱においても，中山らによって 3～4 kHz 付近での顕著なスペクトルピークの存在が報告されている [7, 8]。そこで，2～4 kHz の帯域成分に着目して，STRAIGHT による音響分析を行った。

図–7 に，響きの聴覚印象に関する評定順位が 2 位（No. 1），4 位（No. 4），8 位（No. 9）の音声刺激に関するスペクトル包絡を示す。No. 1 と No. 9 は同一のテノール歌手による歌声と話声，No. 4 は民謡歌手の歌声である。テノール歌手の音声を比較した結果，歌声の方が 2～3 kIIz 付近の成分が最大で 18 dB 強いことが分かった。また，民謡歌手のスペクトル包絡に関しても，3～4 kHz 付近に顕著なピークが観測された。これらの特性は，先行研究における歌唱ホルマントの特性と似ており，とりわけ響きの聴覚印象の順位が高い男性歌手の音声刺激において顕著に観測された。

一方，響きの順位 1 位の女性歌手（メゾソプラノ）の音声を分析した結果，明確な歌唱ホルマントは観測されなかったものの，音源の非周期性指標が 3 kHz 付近において非常に小さいことが明らかとなった。音源の非周期性指標とは，音源波の各帯域における周期成分と非周期成分の割合を示す STRAIGHT の分析パラメータの一つであり，対数振幅スペクトルの上側包絡と下側包絡の差分に対応する。つまり，ある周波数帯域において非周期性が弱いということは，同帯域に強い高調波成分（harmonics）が存在することを示している。その一例として，図–8 に示すように，2～3 kHz 付近に存在するスペクトルピークにおいて，響きの評定順位が 1 位の音声刺激（No. 3）では非周期性指標に顕著な谷が存在するのに対して，順位が低い音声刺激（No. 9）では見られない。



図–8 スペクトル包絡と非周期性指標の比較（左図：No. 3（メゾソプラノ），右図：No. 9）

以上から，2～4 kHz の周波数帯域における歌唱ホルマントと強い高調波成分が，響きの聴覚印象に寄与している可能性が高いことが示された。

6. 歌声らしさの知覚モデルの評価

前章までで，歌声らしさの知覚モデルの層間関係をトップダウン的に調査することで各層の構成要素を抽出した。図–9 に，構築された歌声らしさの知覚モデルを示す。本章では，隣接する層の関係をボトムアップ的に調査することで歌声らしさの知覚モデルの評価を行い，抽出した音響特徴量及び基本的な心理的特徴の妥当性を検証する。

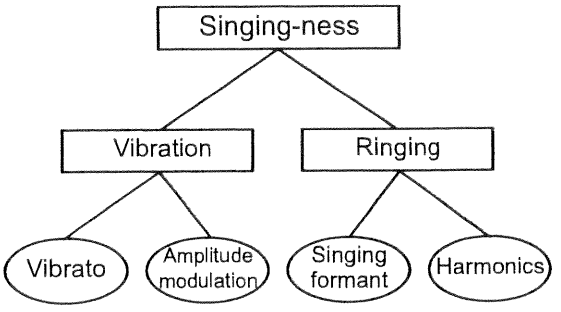
6.1 2–3 層間の評価

第 3 層を構成する各種音響特徴量を操作した合成音を作成し，聴取実験によって第 2 層の各心理的特徴に関する聴覚印象の変化を調査する。

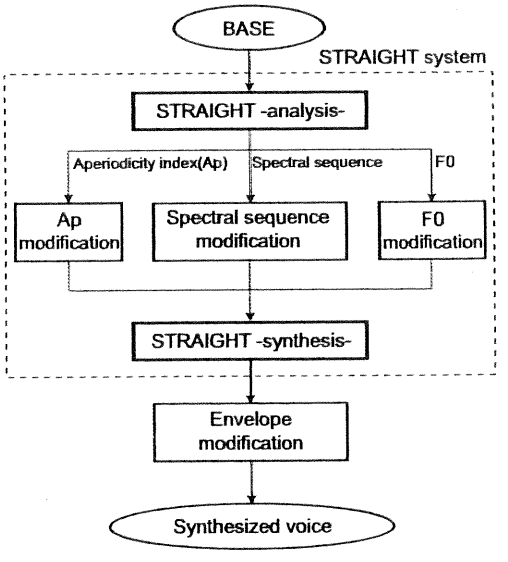
6.1.1 揺れの知覚に関して

F_0 と振幅エンベロープの周期振動成分を操作した合成音を作成し，聴取実験によって揺れに関する聴覚印象を評定した。

合成音は，図–10 に示した合成手順に従って作成した。入力音声の **BASE** とは，3 章の歌声らしさの評定実験の際に多くの聴取者によって話声として知覚されたテノール歌手による通常発話の日本語母音/a/（表–1 の No. 9）を STRAIGHT によって時間伸長したものである。この際，ホルマント定常部 100 ms 区間を 1,000 ms まで伸張し，全体の音韻長は約 1,500 ms となっている。なお，時間伸長による聴覚印象（歌声らしさ，揺れ，響き）への影響の有無を調査するために，上記の定常部を 200 ms 刻みで 1,000 ms まで伸長した 6 種の合成音を作成し，聴取実験によって各聴覚印象の変化を調査した結果，1,000 ms 伸長してもほとんど影響がないことを確認している。



図–9 構築した歌声らしさの知覚モデル



図–10 STRAIGHT を用いた合成音の作成手順

その他の合成音は，**BASE** に各種音響特徴量を付与した以下の 3 種である。

F0-VIB：**BASE** の F_0 パターンに周期振動を付与した合成音

ENV-VIB：**BASE** の振幅エンベロープに周期振動を付与した合成音

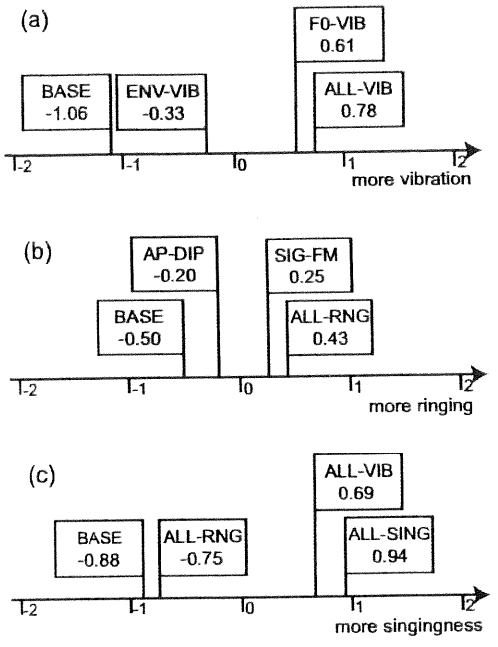
ALL-VIB：**BASE** の F_0 パターンと振幅エンベロープに周期振動を付与した合成音

F0-VIB の F_0 パターン $F0_{\mathrm{vb}}$ の制御は次式のとおりである。

$$F0_{\mathrm{vb}}(t) = (1 + \alpha \sin(2\pi f_{\mathrm{v}} t)) F0_{\mathrm{b}}(t) \qquad (3)$$

ここで，$F0_{\mathrm{b}}$ は BASE の F_0 パターンである。また，α と f_{v} はそれぞれヴィブラートの変動幅と速さ（変調周波数）を制御するパラメータであり，ここでは先行研究 [25] で最も自然なヴィブラート特性と報告されている $F0_{\mathrm{b}}$ の 3%の変動幅で 5.5 Hz の速さで振動するように各パラメータ値を設定した。

ENV-VIB の振幅エンベロープ E_{vb} は，**BASE** の振幅エンベロープ E_{b} に対して上式と同様の方法で作成した。



図–11 各種心理的特徴の聴覚印象に関する刺激音の間隔尺度（(a)：揺れ，(b)：響き，(c)：歌声らしさ）

$$E_{\mathrm{vb}}(t) = (1 + \alpha \sin(2\pi f_{\mathrm{v}} t)) E_{\mathrm{b}}(t) \qquad (4)$$

なお，**BASE** の振幅エンベロープ E_{b} は，次式によって抽出した。

$$E_{\mathrm{b}}(t) = \mathrm{LPF}[|X(t) + j \cdot \mathrm{Hilbert}(X(t))|] \quad (5)$$

ここで，Hilbert$(\cdot)$ はヒルベルト変換である。また，ローパスフィルタ LPF$[\cdot]$ のカットオフ周波数は 30 Hz とした。付与した周期振動の変調周波数は **F0-VIB** と同様の 5.5 Hz とし，振れ幅は 5 章での音響分析の結果を基に，E_{b} の 20%の値に設定した。なお，F_0 及び振幅エンベロープへの操作は，すべて同区間，同位相で付与した。以上すべての合成音の振幅レベルは，最大振幅値で正規化処理を施している。

上記の合成音を実験刺激として，シェッフェの一対比較法（浦の変法）によって揺れの聴覚印象に関する評定実験を行った。なお，評定には揺れに関する 5 段階尺度（+2：先の刺激が非常に揺れている，+1：先の刺激の方が揺れている，0：どちらとも言えない，−1：後の刺激の方が揺れている，−2：後の刺激の方が非常に揺れている）を用いた。聴取者は，正常な聴力を有した大学院生 8 名（男性 7 名，女性 1 名）であり，その他の実験条件・環境は 4.2 節の実験と同様である。

図–11(a) に，揺れの聴覚印象に関する刺激の間隔尺度を示す。いずれの刺激間も，危険率 5%で有意な差があることを確認している。各刺激名の下に記された数値が大きいほど，揺れの聴覚印象が強かったことを表す。この結果から，各音響特徴量を付与することで揺れの聴覚印象が強くなり，とりわけ F_0 の周期変動の影響が大きいことが明らかとなった。

以上から，F_0 と振幅エンベロープの周期変動が，揺れの聴覚印象に寄与していることが確認された。

6.1.2 響きの知覚に関して

2〜4 kHz の周波数帯域におけるスペクトルピーク（歌唱ホルマント）と強い高調波成分を操作した合成音を作成し，聴取実験によって響きの聴覚印象を評定した。

合成音は，図–10 の方法で作成した以下の 3 種である。

SIG-FM：**BASE** に歌唱ホルマントを付与した合成音

AP-DIP：**BASE** に強い高調波成分を付与した合成音

ALL-RNG：**BASE** に歌唱ホルマントと強い高調波成分を付与した合成音

SIG-FM 及び **ALL-RNG** 作成における歌唱ホルマントの制御は，**BASE** の母音定常部 1,000 ms 区間の対数振幅スペクトル包絡に対して，次式で記述される Hanning 窓を用いた荷重関数 W_{sf} を用いて行った。

$$W_{\mathrm{sf}}(f) = \begin{cases} (1 + A_{\mathrm{f}}) \left(1 - \cos\left(2\pi \frac{f}{F_{\mathrm{b}}+1}\right)\right), \\ \left(F_{\mathrm{s}} - \frac{F_{\mathrm{b}}}{2} \le f \le F_{\mathrm{s}} + \frac{F_{\mathrm{b}}}{2}\right) \\ 1, \ (\text{otherwise}) \end{cases} \qquad (6)$$

ここで，F_{s} は **BASE** の対数振幅スペクトル包絡の 3 kHz 付近に存在するホルマント周波数である。また，F_{b} はスペクトル包絡を強調させる帯域幅を，A_{f} は強調させる割合をそれぞれ決定するパラメータである。本実験では，先行研究 [7] で報告されている歌唱ホルマントの特性，及び 5 章の音響分析の結果から，帯域幅を 2 kHz とし，ホルマントピークを 18 dB 持ち上げるように各パラメータを設定し，歌唱ホルマントを付与した。

AP-DIP は，3 kHz 付近に存在するホルマントピークの位置に対応する非周期性指標を関数 $W_{\mathrm{sf}}(f)$ によって弱める（顕著な谷を付与する）処理を施すことで作成した。その際の帯域幅と弱める割合は，5 章の音響分析の結果を基に，歌唱ホルマント制御時と同じ 2 kHz と 18 dB と設定した。この操作によって，3 kHz に存在するホルマント位置に強い高調波成分を付与した。なお，これらすべての合成音の振幅レベルは，最大振幅値で正規化処理を施している。

上記の合成音を実験刺激として，揺れの場合と同様の実験方法・条件で響きに関する聴覚印象について評定した。図–11(b) に響きの聴覚印象に関する刺激の間

隔尺度を示す。いずれの刺激間も，危険率 5%で有意な差があることを確認している。この結果から，各音響特徴量を付与することで響きの聴覚印象が強くなり，とりわけ歌唱ホルマントの影響が大きいことが明らかとなった。

以上から，スペクトル包絡における 3 kHz 付近の顕著なピークと，同じく 3 kHz 付近での強い高調波成分が，響きの聴覚印象に寄与することが確認された。

6.2 1–2 層間の評価

4 章で構築した第 1 層と第 2 層の関係について検証する。前節で揺れと響きの聴覚印象に寄与することが示された各種音響特徴量を操作した合成音を作成し，聴取実験によって歌声らしさの聴覚印象を評定する。これにより，2 層を構成する心理的特徴の妥当性を検証する。

実験に用いた合成音は，前節と同じ手法を用いて **BASE** に対して各種音響特徴量を付与した以下の 4 種である。

ALL-VIB：**BASE** の F_0 パターンと振幅エンベロープに周期振動を付与した合成音

ALL-RNG：**BASE** に歌唱ホルマントと強い高調波成分を付与した合成音

ALL-SING：**BASE** にすべての音響特徴量を付与した合成音

なお，合成の際に設定した各種パラメータは，すべて前節と同じである。

上記の合成音を実験刺激として，シェッフェの一対比較法（浦の変法）によって歌声らしさの聴覚印象に関する評定実験を行った。なお，評定には歌声らしさに関する 5 段階尺度（+2：先の刺激が非常に歌声らしい，+1：先の刺激の方が歌声らしい，0：どちらとも言えない，−1：後の刺激の方が歌声らしい，−2：後の刺激の方が非常に歌声らしい）を用いた。

図–11(c) に歌声らしさ聴覚印象に関する刺激の間隔尺度を示す。いずれの刺激間も，危険率 5%で有意な差があることを確認している。**BASE** に各音響特徴量を付与することで，歌声らさしさの評定値が向上する結果となった。これは，それぞれの心理的特徴が，歌声らしさの聴覚印象に寄与することを示している。また，**ALL-VIB** が **ALL-RNG** に比べてより歌声らしいと知覚されている結果となった。これは，歌声らしさの聴覚印象には，揺れの心理的特徴の方が強く寄与していることを示している。ただし，**All-SING** が **ALL-VIB** よりも評定値が大きいことから，響きの心理的特徴は，揺れと共存することで歌声らしさの聴覚印象に寄与すると考えられる。

以上から，前章までで抽出した歌声らしさの知覚モデルの各層の構成要素の妥当性が検証された。これにより，歌声らしさの聴覚印象に影響を与える歌声特有の音響特徴量として，F_0 の準周期振動成分であるヴィブラートとそれに伴った振幅エンベロープの周期振動，更には歌唱ホルマントと同帯域の強い高調波成分の存在が明らかとなった。

7. ま と め

本論文では，歌声知覚と音響特徴量の関係を調査する新たな方法として，歌声らしさの知覚モデルを提案した。このモデルは，歌声らしさの聴覚印象（第 1 層）と音響特徴量（第 3 層）の関係を基本的な心理的特徴（第 2 層）を介して記述した多層構造モデルであり，本論文では各層間の関係を調査することで歌声知覚と音響特徴量の関係を明らかにした。

第 1 層と 2 層の関係については，多様な歌唱法の歌声・話声が混在する音声データを対象にした聴取実験を行い，MDS による歌声らしさの空間を構築した。そして，歌声らしさの知覚に寄与する表現語の選出，及びその表現語に関する歌声らしさの空間における重回帰分析を行なった。その結果，揺れ，響き，明瞭さの知覚が歌声らしさの知覚に大きな影響を与えていることが明らかとなった。その中でも，揺れと響きは歌声と話声を聞き分ける上で重要な心理的特徴であることが分かった。

第 2・3 層間の関係では，揺れと響きの聴覚印象に寄与する音響的特徴を調査した。その結果，揺れの聴覚印象が強い音声ほど，ヴィブラートと呼ばれる F_0 の準周期的振動成分（変調周波数 4〜6 Hz 程度）が含まれており，更には振幅エンベロープが，ヴィブラート振動に同期して振動していることが明らかとなった。一方，響きに関しては，歌唱ホルマントと呼ばれる 3 kHz 付近の顕著なスペクトルピーク成分と，同帯域における強い高調波成分が重要であることが示された。

最後に，構築したモデルの評価として，第 3 層を構成する各種音響特徴量を操作した合成音を作成し，揺れ・響き，更には歌声らしさの聴覚印象への影響を調査した。その結果，歌声らしさの知覚には，揺れに関する聴覚印象及びそれに起因する音響特徴量の影響が大きいことが示された。また，響きの聴覚印象は，揺れの聴覚印象と共存することで歌声らしさの知覚に寄与することが示された。

以上のように，歌声らしさの知覚モデルに基づき，各層間の対応関係と各層の構成要素を調査することで，高次の心理的特徴である歌声らしさの聴覚印象に寄与する歌声特有の音響的特徴が明らかになった。いずれの音響特徴量も先行研究で報告されているが，各特徴

の歌声知覚にける役割とその重要性を明確にしたことは，本研究の大きな成果と言える。この成果は，歌声らしさの知覚モデルに基づいた新たな研究手法の有効性を示すだけでなく，高品質の歌声合成や歌声知覚・生成機構の解明に大きく貢献できると考えられる。

謝　辞

本研究を進めるにあたり，北海道医療大学の榊原健一氏に数々の貴重なご助言をいただいた。本研究の一部は，科学研究費補助金（13610079）及び総務省戦略的情報通信研究開発推進制度SCOPE（071705001）の援助を受けて行われた。

文　献

[1] C.E. Seahore, "Phychology of the vibrato in voice and instrument," *Studies in the Psychology of Music*, Vol. 1 (The University Press, Iowa, 1932).
[2] J. Hakes, T. Shipp and T. Doherty, "Acoustic characteristics of vocal oscillations: Vibrato exaggerated vibrato, trill, and tillo," *J. Voice*, 1, 326–331 (1987).
[3] Y. Horii, "Acoustic analysis of vocal vibrato: A theoretical intepretation of data," *J. Voice*, 3, 36–43 (1989).
[4] G. Krom and G. Bloothooft, "Timing and accuracy of fundamental frequency changes in singing," *Proc. ICPhS 95*, Stockholm, Vol. I, pp. 206–209 (1995).
[5] H. Mori, W. Odagiri and H. Kasuya, "F0 dynamics in singing: Evidence from the data of a baritone singer," *IEICE Trans. Inf. Syst.*, **E87-D**, 1086–1092 (2004).
[6] M. Akagi, M. Iwaki and T. Kitakaze, "Fundamental frequency fluctuation in continuous vowel utterance and its perception," *Proc. ICSLP 98*, Sydney, Vol. 4, pp. 1519–1522 (1998).
[7] J. Sundberg, "Articulatory intepretation of the "singing formant"," *J. Acoust. Soc. Am.*, 55, 838–844 (1974).
[8] S. Wang, "Singer's high formant associated with different larynx position in styles of singing," *J. Acoust. Soc. Jpn. (E)*, 7, 303–314 (1986).
[9] W.T. Bartholomew, "A physical definition of "good voice-quality" in the male voice," *J. Acoust. Soc. Am.*, 6, 25–33 (1934).
[10] I. Nakayama, "Comparative studies on vocal expressions in Japanese traditional and Western classical-style singing using common verse," *Proc. ICA 2004*, Mo4, C1.1 (2004).
[11] 小林範子, 東倉洋一, 天白成一, 新美成二, "日本の伝統歌唱における生成面の特徴," 音響学会音声研資, SP-89-147, pp. 39–45 (1990).
[12] 西内美登里, 大串健吾, "専門家と非専門家の歌声の評価," 音響学会聴覚研資, H-90-1, pp. 1–6 (1990).
[13] 上田和夫, "音色の表現語に階層構造は存在するか," 音響学会誌, 44, 102–107 (1988).
[14] 木戸　博, 粕谷英樹, "通常発話の声質に関連した日常表現語—聴取実験による抽出—," 音響学会誌, 57, 337–344 (2001).
[15] H. Kawahara, I. Matsuda-Katsuse and A. Cheveigne, "Restructuring speech representationsusing a pitch adaptive time-frequency smoothing and an instantaneous-frequency based on F0 extraction: Possible role of arepetitive structure in sounds," *Speech Commun.*, 27, 187–207 (1999).
[16] T. Saitou, M. Unoki and M. Akagi, "Development of an F0 control model based on dynamic characteristics for singing-voice synthesis," *Speech Commun.*, 46, 405–417 (2005).
[17] N. Minematsu, B. Matsuoka and K. Hirose, "Prosodic modeling of nagauta singing and its evaluation," *Proc. Seech Prosody 2004*, pp. 487–490 (2004).
[18] 河原英紀, 生駒太一, 森勢将雄, 高橋　徹, 豊田健一, 片寄晴弘, "歌唱音声モーフィングに基づく声質と歌い回し転写の知覚的検討," インタラクション 2007 論文集, pp. 113–120 (2007).
[19] 中山一朗, "日本語を歌・唄・謡う" 音響学会誌, 59, 688–693 (2003).
[20] 天坂格朗, 長沢伸也, "官能評価の基礎と応用," 日本規格協会 (2000).
[21] 桑原尚夫, 大串健吾, "アナウンサー音声の音響的特徴," 信学論, **J66-A**, 545–552 (1983).
[22] 林知己夫, 飽戸　弘, "多次元尺度解析法" (サイエンス社, 東京, 1996).
[23] 齋藤　毅, 鵜木祐史, 赤木正人, "歌声の F0 制御モデルにおけるパラメータ決定に関する考察," 音響学会聴覚研資, H-2003-11, pp. 653–658 (2003).
[24] W. Vennard, *Singing: The Mechanism and the Technic*, 2nd ed. (Fischer, New York, 1967).
[25] 齋藤　毅, 鵜木祐史, 赤木正人, "自然性の高い歌声合成のためのヴィブラート変調周波数の制御法の検討," 信学技報, TL2005-10, pp. 13–17 (2005).