社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2008－MUS－74 (5)
2008－SLP－70 (5)
2008／2／8

SingBySpeaking:
歌声知覚に重要な音響特徴を制御して話声を歌声に変換するシステム

齋藤 毅 † 後藤 真孝 † 鵜木 祐史 †† 赤木 正人 ††

† 産業技術総合研究所
†† 北陸先端科学技術大学院大学 情報科学研究科

あらまし 本稿では，歌詞の朗読音声（話声）を歌声に変換する歌声合成システム SingBySpeaking について述べる．このシステムは，音声分析合成系 STRAIGHT による分析/合成処理過程において，基本周波数（F0），スペクトル，音韻長を制御するモデルによって歌声特有の音響特徴を操作することで話声を歌声に変換する．F0 制御モデルは，楽譜情報から得られるメロディの遷移の概形に対して，4 種類の動的変動成分（オーバーシュート，ヴィブラート，プレパレーション，微細変動）を付与することで歌声の F0 変化パターンを生成する．スペクトル制御モデルは，話声のスペクトルに対して，歌唱ホルマントとヴィブラートに同期したホルマントの振幅変調を付与することで歌声のスペクトル形状を生成する．音韻長制御モデルは，楽曲のテンポに基づいて，話声中の各音韻長を歌声の音韻長に伸長する．システムで合成された音声を聴取実験によって評価した結果，各種音響特徴を制御することで話声から歌声に変換され，すべての特徴を制御した合成音の音質は原音声と同程度であることを示した．

SingBySpeaking:
Singing Voice Conversion System from Speaking Voice By Controlling Acoustic Features Affecting Singing Voice Perception

Takeshi Saitou † Masataka Goto † Masashi Unoki †† Masato Akagi ††

† National Institute of Advanced Industrial Science and Technology (AIST)
†† School of Information Science, Japan Advanced Institute of Science and Technology

Abstract This paper describes a novel singing voice synthesis system *SingBySpeaking* that can synthesize a singing voice, given a speaking voice reading the lyrics of a song and its musical score. The system is based on the speech manipulation system *STRAIGHT* and comprises three models controlling three acoustic features unique to singing voices: the fundamental frequency (F0), phoneme duration, and spectrum. Given the musical score and its tempo, the F0 control model generates the F0 contour of the singing voice by controlling four types of F0 fluctuations: overshoot, vibrato, preparation, and fine fluctuation. The duration control model lengthens the duration of each phoneme in the speaking voice by considering the duration of its musical note. The spectral control model converts the spectral envelope of the speaking voice into that of the singing voice by controlling both the singing formant and the amplitude modulation of formants in synchronization with vibrato. Experimental results show that the proposed system can convert speaking voices into singing voices whose naturalness is almost the same as actual singing voices.

1 はじめに

本稿では，歌声特有の音響特徴制御することで話声を歌声に変換する歌声合成システム SingBySpeaking を提案する．歌を歌うことは，音楽を楽しむ最も手近な手段であると同時に，歌詞である言語情報に加え感情や想いといった非言語情報を表出するための重要な手段である．その為，歌声合成システムの構築は，計算機による音楽の新たな楽しみ方を創造するだけでなく，人間の音声コミュニケーションを理解する上でも重要な取り組みである．

現在の歌声合成の研究は，テキスト又は歌詞から歌声を合成する *text-to-singing (lyrics-to-singing) synthesis* のアプローチによる取り組みが主流であ

る[1, 2, 3, 4]．これらは，話声を対象とした *text-to-speech synthesis* で用いられる波形接続合成やHMM 合成といったコーパスベースの合成手法に基づいた実用性の高いものが多く，特にYAMAHAの VOCALOID[2] は市販の歌声合成ソフトウェアとして計算機音楽の新しい可能性を示している．

それに対して，我々は，話声から歌声を合成する *speech-to-singing synthesis* という新しいアプローチで歌声合成システムを構築することを目指す．このシステムは，歌声特有の音響特徴に着目し，音声分析・合成系（ボコーダ）による処理過程においてそれらの音響特徴を制御することによって歌詞の朗読音声を歌声へ変換するものである．我々は，この方法を用いることで，従来の歌声合成システム以上の自然な歌声の合成と，「歌詞を朗読さえすれば元の声質を保持した歌声を生成できる」という新たな歌声アプリケーションが実現できると考えている．更には，歌声特有の各種音響特徴を操作・変換した歌声を合成できるシステムの枠組み自体が，歌声の知覚・生成機構を解明する有効な手法になり得ると考えている．

歌声特有の音響特徴，更にはそれら特徴の歌声知覚に与える影響に関しては，これまでに数多く調査されてきている[5, 6]．しかし，いずれの研究も，特定の歌唱法や歌唱者による限定された歌声を対象にしたものであり，話声と歌声の違いという観点からの取り組みは少ない．その中で，大石ら[7]による話声と歌声の自動識別システムや，辻ら[8]による“歌声らしい声”の心理的，音響的な分析においては，話声と歌声の音響構造の違い，更にはその違いが知覚に与える影響について調査している．また，筆者らの研究[9]において，F0，スペクトル，音韻長の各音響パラメータに話声と歌声の違いを規定する音響特徴が存在することを確認している．

そこで，本稿では，F0，スペクトル，音韻長の各音響パラメータにおける歌声特有の音響特徴を制御するモデルを構築し，それらのモデルを音声分析合成系 STRAIGHT[10] の処理体系に組み込むことで，話声から歌声に変換する歌声合成システムを提案する．第 2 章では，提案する歌声合成システムの概要を述べる．第3~5 章までは，F0，音韻長，スペクトルの各制御モデルについて概説する．特に第 3 及び 5 章では，歌声の F0 とスペクトル特有の音響特徴が歌声知覚に与える影響についても示す．そして，第 6 章では本システムによる歌声合成を行い，合成音の評価，更には話声から歌声に変化する上で各種音響特徴が担う役割についても述べる．最後に，第 7 章において本歌声システムの構築を通じて明らかになったことを述べ，今後の研究展望を示す．

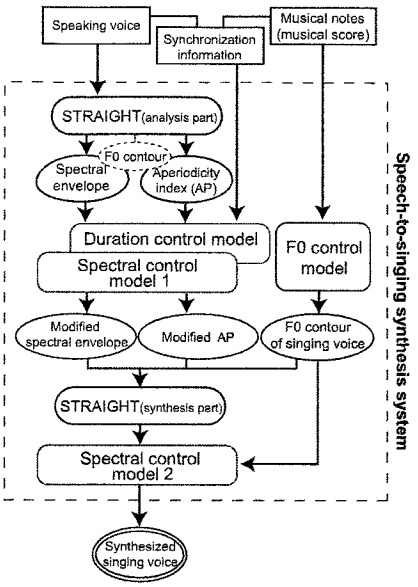

Fig. 1 *Block diagram of the proposed singing voice synthesis system.*

2 歌声合成システムの処理体系

図 1 に歌声合成システム SingBySpeaking の概要を示す．システムの入力は，合成したい歌の歌詞の朗読音声（speaking voice），その歌の譜面情報（musical score），そして朗読音声中の音韻（または単語）と譜面中の音符の対応関係を記述した情報（synchronization information）の 3 つである．尚，朗読音声のセグメンテーションと，セグメンテーションされた各音韻と音符との対応付けは手動で行う必要がある．このシステムでは，以下の 6 つの手続きによって歌声合成音が生成される．

1. 朗読音声を STRAIGHT によって F0 変化パターン，スペクトル系列，非周期性指標系列の音響パラメータに分解する
2. 歌声 F0 制御モデルによって譜面情報から歌声の F0 変化パターンを生成する
3. 音韻長制御モデルによって朗読音声中の各音韻のスペクトルと非周期性指標の時間系列を伸長する
4. スペクトル制御モデル 1 によって時間伸長後の母音区間のスペクトル包絡と非周期性指標

を加工する

5. 生成・加工した各音響パラメータを用いてSTRAIGHTによって歌声を合成する
6. スペクトル制御モデル2によって合成歌声の振幅エンベロープを加工する

3 F0制御モデル

歌声のF0制御を行うには，F0変化パターンに含まれる特徴を抽出し，それらを制御できるモデルが必要となる．本章では，歌声のF0変化特有の音響特徴であるF0動的変動成分について述べ，それら特徴を制御可能な歌声F0制御モデルを概説する．そして，F0動的変動成分が歌声知覚に与える影響について示す．

3.1 歌声特有のF0動的変動成分

筆者らの先行研究[11]において様々な歌声データのF0変化パターンを分析した結果，以下に示す4種類のF0動的変動成分が歌唱スタイルや歌唱者に関係なく存在することが明らかになっている．

オーバーシュート (Overshoot) ：滑らかな音高の変化，およびその直後に目的音高を越える瞬時的な変動成分

ヴィブラート (Vibrato) ：同一音高区間で観測される4~8 Hzの準周期的な変動成分

プレパレーション (Preparation) ：音高変化直前に変化とは逆方向振れる瞬時的な変動成分

微細変動 (Fine fluctuation) ：発声区間全体に観測される不規則で細かい変動成分

図2に，アマチュア歌手による日本童謡「七つの子」歌唱時のF0変化パターンと，そこに含まれるF0動的変動成分を示す．オーバーシュートとプレパレーションは，歌唱特有のメロディに基づく急峻なF0遷移の結果として生起すると考えられ，その特性は歌唱技量に差がある歌唱者間でも大きな違いが無いことが確認されている[12]．ヴィブラートは，特に西洋オペラ歌唱において音色を豊かにする要素と考えられており[13]，その特性は歌唱技量の差によって明確に異なる事が分かっている[14]．微細変動は，上記3種の成分を取り除いた後に残る不規則的な変動であり，その特性は変調周波数が15~20 Hz程度，偏移幅が約±20 centと報告されている[15]．

3.2 F0制御モデル

図3に歌声のF0制御モデルの概要を示す．このモデルは，譜面中の各音符をステップ関数で記述し，

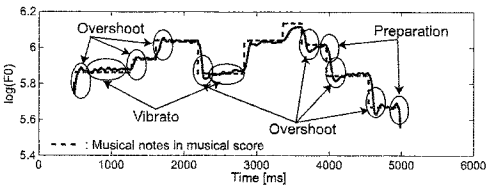

Fig. 2 *Examples of F0 fluctuations in the singing voice of an amateur singer.*

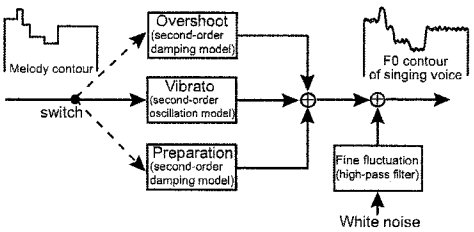

Fig. 3 *Block diagram of the F0 control model for singing voices.*

それらを重ね合わせることで生成したメロディの遷移の概形（Melody contour）に対して，4種のF0動的変動成分を制御・付与することによって歌声のF0変化パターンを生成する．

オーバーシュート，ヴィブラート，プレパレーションは，メロディの遷移の概形を複数のフィルタに通すことで制御される．フィルタは，次式の二次系伝達関数のインパルス応答で与えられる．

$$H(s) = \frac{k}{s^2 + 2\zeta\omega s + \omega^2}, \tag{1}$$

ここで，ωは固有角周波数，ζは減衰項，kは振幅項である．インパルス応答$h(t)$は，ζの値に従って以下のように与えられる．

$$h(t) = \begin{cases} \frac{k}{2\sqrt{\zeta^2-1}}(\exp(\lambda_1\omega t) - \exp(\lambda_2\omega t)), & |\zeta| > 1 \\ \frac{k}{\sqrt{1-\zeta^2}}\exp(-\zeta\omega t)\sin(\sqrt{1-\zeta^2}\omega t), & 0 < |\zeta| < 1 \\ kt\exp(-\omega t), & |\zeta| = 1 \\ \frac{k}{\omega}\sin(\omega t), & |\zeta| = 0 \end{cases} \tag{2}$$

ここで，$\lambda_1 = -\zeta + \sqrt{\zeta^2 - 1}$，$\lambda_2 = -\zeta - \sqrt{\zeta^2 - 1}$である．そして，オーバーシュートとプレパレーションは減衰振動モデル（$0 < |\zeta| < 1$），ヴィブラートは定常振動モデル（$|\zeta| = 0$）で記述される．また，各F0動的変動成分の特性は，パラメータω，ζ，kによって制御される．オーバーシュートとプレパレーションに関してはζによって目的音高を越える大きさ，ωとζによって変動の持続時間がそれぞれ制御される．ヴィブラートに関しては，ωによって振動の速さ（vibrato rate），kによって

Table 1 parameter values for controlling F0 fluctuations.

F0 fluctuation	Ω[rad/ms]	ζ	k
overshoot	0.0348	0.542	0.0348
vibrato	0.0345	0	0.0018
fine fluctuation	0.0292	0.668	0.0292

振動の大きさ（vibrato extent）が制御される．本稿では，各制御パラメータを表1の値に設定した．これらの値は，歌声データベース「日本語を歌・唄・謡う」[16] 中の歌声データからSTRAIGHTによって抽出したF0とF0制御モデルで生成されるF0の誤差が最少となるように，非線形最少自乗法[17] によって決定した．

微細変動は，白色雑音をカットオフ周波数10 Hzの低域通過フィルタに通した後，最大振幅が5 Hzになるように正規化することで生成し，F0変化パターン全体に付与する．尚，F0制御モデルの詳細に関しては文献[11] を参照されたい．

3.3 F0動的変動成分の知覚への影響

4種のF0動的変動成分が歌声知覚に与える影響は，メロディ変化に個々のF0動的変動成分を付与した合成音を作成し，それらを聴取実験によって評価することで調査した．

合成音は，女性3名が母音/a/のみで歌唱した日本童謡「七つの子」を対象に，STRAIGHTと歌声F0制御モデルによって作成した以下の5種である．

NORMAL：STRAIGHTによる分析・再合成音．（音質は原音声とほぼ同じ）

SYN-BASE：F0動的変動成分は付与せず，メロディ変化のみの合成音

SYN-OUS：**SYN-BASE**にオーバーシュート成分のみ制御した合成音

SYN-VB：**SYN-BASE**にヴィブラート・微細変動成分のみ制御した合成音

SYN-PRE：**SYN-BASE**にプレパレーションのみ制御した合成音

SYN-ALL：**SYN-BASE**にすべてのF0動的変動成分を制御した合成音

上記の合成音を実験刺激とし，シェッフェの一対比較法[18] によって，歌声の自然性に関する間隔尺度を求めた．実験に用いた評価尺度は，歌声の自然性に関する7段階評価（+3:先の刺激がとても自然，+2:自然，+1:やや自然，0:どちらとも言えない，-3:後の刺激がとても自然，-2:自然，-1:やや自然）である．被験者は，正常な聴力を有した大学院生6名（男性5名，女性1名）である．実験環境は防音室内で行い，刺激音はヘッドホン(STAX SR-404) を介して呈示した．

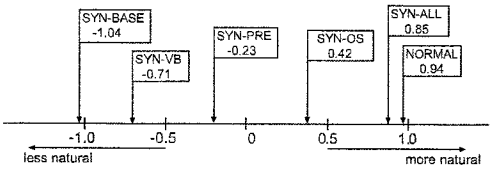

Fig. 4 *Perceptual effects of four types of F0 fluctuation.*

実験結果を図4に示す．図の水平軸は，各刺激の歌声としての自然性を表し，右に布置されている刺激ほど，自然な歌声として知覚されたことを示す．この結果から，各F0動的変動成分を付与することで歌声の自然性が増加し，中でもオーバーシュートの影響が大きいことが分かる．以上から，各F0動的変動成分が，歌声を知覚する上で重要な役割を担っていることが確認できた．

4 音韻長制御モデル

歌声中の個々の音韻長は，各音韻に割り当てられた音符と楽曲のテンポによって決定される．その為，歌詞を朗読した場合の音韻長とは大きく異なる．本章では，個々の音符が割り当てられた音韻（もしくは単語）ごとに時間伸長処理を行い，話声の時間構造を歌声に変換する音韻長制御モデルについて概説する．

図5に音韻長制御モデルの概要を示す．このモデルは，話声の音声波形ではなく，STRAIGTHによって得られたスペクトルと非周期性指標の時間系列を線形補間することで時間伸長を行う．また，各音韻長を音符長に従って一様に伸長するのではなく，手動で与えられる各音韻の子音–母音境界を子音部＋結合部＋母音部に自動セグメンテーションし，各部に対して伸長処理を行う．尚，結合部分は子音–母音境界の-10 ～ 30 msまでの計40 msとし，この長さはすべての音韻において一律としている．

子音部の時間長は，予め朗読音声中の子音長に対する歌声中の子音長の比率を求めておき，その比率に従って伸長処理を行う．日本童謡「七つの子」を女性3名，男性1名がそれぞれ朗読/歌唱したデータから伸長比率を算出した結果，同じ調音

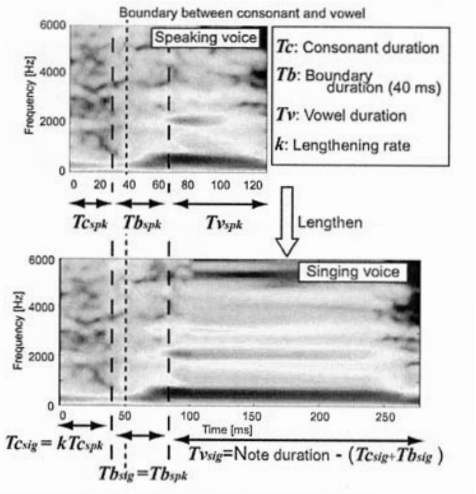

Fig. 5 *Schema of the duration control model.*

様式の子音において類似していることが確認でき，その値は，摩擦音で1.58，破裂音で1.13，半母音で2.07，鼻音で1.77，そして/y/で1.13と設定した．尚，これらの値は，割り当てられた音符の種類や歌唱者の性別に依存せず，ほぼ一定であることを確認している．

結合部は，時間伸長を行わない．つまり，結合部の長さは40 msで固定とする．

母音部は，伸長の対象としている音韻に割り当てられた音符長から，伸長した子音部の長さと結合部の40 msを差し引いた時間長に伸長する（図5を参照）．

5 スペクトル制御モデル

歌声のスペクトル制御を行うには，歌声のスペクトル特有の音響特徴（以後スペクトル特性と呼ぶ）を抽出し，それらを制御できるモデルが必要となる．本章では，2種の歌声特有のスペクトル特性を示し，それら特徴を制御可能なスペクトル制御モデルについて概説する．そして，スペクトル特性が歌声知覚に与える影響についても示す．

5.1 歌声特有のスペクトル特性

歌声特有のスペクトル特性の一つに，Sundberg[19]が“歌唱ホルマント（singer's formant）”と命名した男性のオペラ歌唱の3 kHz付近において観測される顕著なホルマントピークがある．また，Nakayama[20]や小林ら[21]によって，長唄や民謡といった邦楽的歌唱においても同様のスペクトルピークの存在が報告されている．更に，Wang[22]

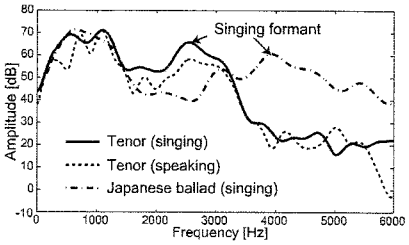

Fig. 6 *Examples of singing formant near 3 kHz.*

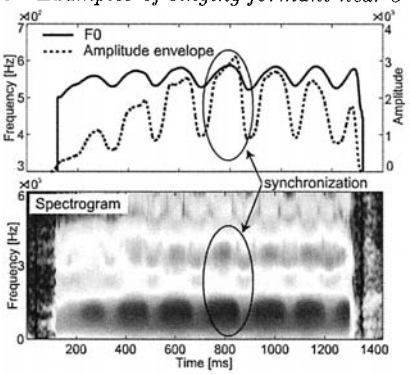

Fig. 7 *Example of formant amplitude modulation (AM) in synchronization with vibrato of the F0.*

や筆者ら[9]の研究によって，歌唱ホルマントが歌唱スタイルや性別に依存せず，多くの歌声に共通して存在することが明らかになってきている．図6は，オペラ歌唱と，長唄歌唱における歌唱ホルマントの一例を示したものであり，話声の同帯域におけるホルマントピークに比べ，12~18 dB強いことが分かっている．歌唱ホルマントの生成は，喉頭全体を下げた発声によって生じる広い下咽頭と比較的狭い喉頭官で形成される声道形状に起因している[23]と考えられ，聴覚的には声に響きや明瞭さを与えると言われている[24]．

また，歌声特有のF0変化に連動し，スペクトル構造も動的変動することが知られている．その代表的な変動として，音声振幅全体がヴィブラートによって振幅変調し，それによって個々のホルマントも振幅変調することが報告されている[25]．これは，F0変化中にヴィブラートが存在する場合は必然的に存在する特徴であり，歌声固有の音響特徴と言える．図7に，オペラ歌唱による母音/a/発声時のサウンドスペクトログラム，F0変化パターン，振幅エンベロープを示す．振幅エンベロープとホルマントが，F0変化中のヴィブラートに同期して振幅変調していることが確認できる．

5.2 スペクトル制御モデル

図1に示すように，スペクトル制御モデルは，2つの手続きから構成され，1番目のモデルで歌唱ホルマントが，2番目のモデルでヴィブラートに同期した音声振幅及びホルマントの振幅変調が制御される．

図8に，スペクトル制御モデル1の概要を示す．このモデルは，次式に従って，話声の母音区間のスペクトルにおいて3 kHz付近に存在するピークを強調することで歌唱ホルマントを制御・付与する．

$$S_{sg}(f) = W_{sf}(f) S_{sp}(f), \tag{3}$$

ここで，$S_{sp}(f)$ と $S_{sg}(f)$ は，それぞれ話声と歌声のスペクトルである．$W_{sf}(f)$ は，話声のスペクトルピークを強調させる荷重関数で次式で表わされる．

$$W_{sf}(f) = \begin{cases} (1+k_{sf})(1-\cos(2\pi \frac{f}{F_b+1})), & |f-F_s| \leq \frac{F_b}{2} \\ 1, & \text{otherwise} \end{cases} \tag{4}$$

ここで，F_s は，$S_{sp}(f)$ の3 kHz付近のピーク周波数で，F_b は強調させる帯域幅を，そして k_{sf} は強調させる割合をそれぞれ調整するパラメータである．本稿では，F_b を2000 Hzに固定し，k_{sf} は F_s のピークを12 dB強調させるように設定した．また，非周期性指標に関しては，3 kHz付近に存在するディップを関数 $W_{sf}(f)$ によって強める（顕著な谷を付与する）処理を行う．

図9にスペクトル制御モデル2の概要を示す．このモデルは，スペクトル制御モデル1の処理後に合成された歌声の振幅エンベロープに対して振幅変調を付与することで，音声振幅とホルマントの振幅変調を制御する．この振幅変調は，F0制御モデルによってヴィブラートが付与された区間において，次式によって付与される．

$$E_{sg}(t) = (1 + k_{am} \sin(2\pi f_{am} t)) E_{sp}(t), \tag{5}$$

ここで，$E_{sp}(f)$ と $E_{sg}(f)$ は，それぞれ話声と歌声の振幅エンベロープである．f_{am} と k_{am} は，振幅変調の速さ（rate）の大きさ（extent）をそれぞれ制御するパラメータである．本稿では，f_{am} を5.5 Hz，k_{am} を0.2にそれぞれ設定した．

5.3 スペクトル特性の知覚への影響

2種のスペクトル特性が歌声知覚に与える影響を調査するために，持続発話母音/a/ に個々の特徴を付与した合成音を作成し，聴取実験によって評価した．

合成音は，STRAIGTHと歌声F0制御モデルによって作成した以下の5種である．

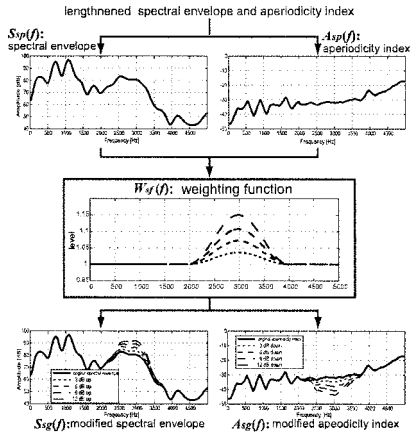

Fig. 8 *Schema of the spectral control model 1.*

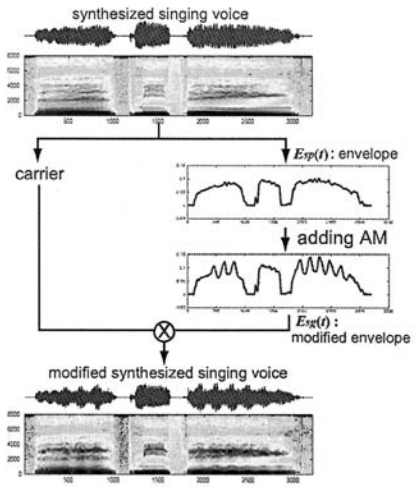

Fig. 9 *Schema of the spectral control model 2.*

BASE：男性による日本語母音/a/の持続発話音声（時間長は150 ms，平均F0は179 Hz）

SYN-SF：**BASE**に歌唱ホルマントを付与した合成音

SYN-AM：**BASE**に振幅及びホルマントの振幅変調を付与した合成音

SYN-ALL：**BASE**にすべての特徴を付与した合成音

上記の合成音を実験刺激として，3.3節と同様の方法・条件で聴取実験を行った．

実験結果を図10に示す．この結果から，各種音響特徴を**BASE**に付与することで歌声の自然性が増加し，中でも音声振幅とホルマントの振幅変

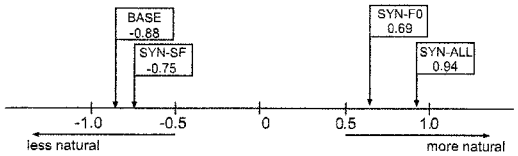

Fig. 10 *Perceptual effects of two types of spectral characteristics.*

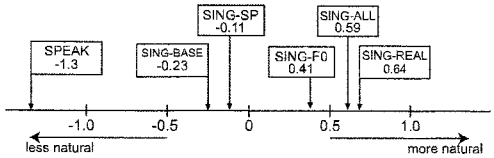

Fig. 11 *Perceptual effects of acoustic features unique to singing voices.*

調の影響が大きいことが分かる．一方で，歌唱ホルマントのみを制御した場合の影響は小さく，振幅変調と共存することで歌声の自然性に寄与することが明らかとなった．

6 歌声合成音の評価

本章では，提案した歌声合成システムを用いて，前章までで述べてきたF0，スペクトル，音韻長における各種音響特徴を個々に制御した歌声合成音を作成し，それらを聴取実験によって評価する．これにより，提案している歌声合成音のシステム評価を行うと同時に，各種音響特徴の歌声知覚への影響を比較する．

6.1 歌声合成

合成音は，男女各1名が日本童謡「七つの子」の歌いだし“からすなぜなくの”を朗読した音声を対象に作成した以下の6種である．

SPEAK：歌詞の朗読音声

SING-BASE：音韻長制御を行った合成音（F0はメロディの遷移の概形を使用）

SING-F0：音韻長制御とF0制御を行った合成音

SING-SP：音韻長制御とスペクトル制御1，2を行った合成音

SING-ALL：すべての制御を行った合成音

SING-REAL：**SPEAK**と同じ人間による歌声

尚，**SPEAK**と**SING-REAL**はSTRAIGHTによって分析・再合成されたものを使用した．

6.2 聴取実験

上記合成音を実験刺激として，3.3節と同様の方法・条件で聴取実験を行った結果を図11に示す．

この結果から，各特徴を付与することで自然な歌声として知覚されるようになり，すべての音響特徴を付与した合成音の自然性は，原歌声と同程度であることが確認された．各種特徴の影響を比較してみると，音韻長制御とメロディの遷移の概形を付与しても（合成音**SING-BASE**）自然な歌声とは知覚されず，それに加えてF0とスペクトルの音響特徴が歌声を知覚する上で必要であることが分かる．その中で，F0動的変動成分の影響はとりわけ大きく，スペクトル特性はF0動的変動成分と共存することで強い影響を与えていることが示された．また，被験者の内観報告から，歌声合成の声質が話声からほとんど変化していないことも明らかとなった．

7 まとめ

本稿では，歌詞の朗読音声を歌声に変換する歌声合成システムSingBySpeakingを提案した．このシステムは，F0，音韻長，スペクトルをそれぞれ制御するモデルで構成され，各モデルにおいて歌声特有の音響特徴を朗読音声に制御・付与することで歌声を合成する．F0制御モデルは，譜面情報から得られるメロディの遷移の概形に対して4種のF0動的変動成分（オーバーシュート，ヴィブラート，プレパレーション，微細変動）を付与することで歌声のF0変化パターンを生成する．音韻長制御モデルは，音符が割り当てられた音韻毎に，子音部，結合部，母音部それぞれに対して楽曲のテンポと音符の種類に従って時間伸長し，話声から歌声の音韻長に変換する．スペクトル制御モデルは，歌唱ホルマントとヴィブラートに同期した音声振幅及びホルマントの振幅変調を話声に付与することで，歌声のスペクトルを生成する．

システムで合成された歌声を評価した結果，実際の歌声と同程度の音質で，且つ話声の声質を崩すことなく自然な歌声を合成することが確認できた．更に，話声から歌声に変換するには，メロディ情報の付与とそれに伴う音韻長制御では不十分であり，F0とスペクトルにおける歌声特有の音響特徴，とりわけF0動的変動成分が重要な役割を果たしていることが明らかとなった．これらの結果は，*speech-to-singing synthesis*とい新しいアプローチで歌声合成に取り組んだ結果生み出されたものであり，計算機音楽における新しい歌声アプリケーションに応用できるだけでなく，歌声特有の知覚・生成機構を解明する上での有用な知見となる．

今後は，歌詞の朗読音声のセグメンテーションを自動化することで，誰でも簡単に利用できる歌声合成システムへ発展させることが必要である．また，本システムを用いた歌声合成・知覚実験の枠組みによって，新たな歌声特有の音響特徴や，歌唱者や歌唱スタイルの違いを規定する音響特徴の抽出を行う予定である．

謝辞 本研究の一部は，科学技術振興機構 CrestMuse プロジェクトによる支援を受けた．最後に，有益なコメントを頂いた榊原健一氏（北海道医療大学）に感謝する．

参考文献

1) J. Bonada and X. Serra, "Synthesis of the Singing Voice by Performance Sampling and Spectral Models," IEEE Signal Processing Magazine, Vol. 24, Iss.2, pp. 67-79, 2007.

2) 剣持秀紀，大下隼人，"歌声合成システム VOCALOID," 情報処理学会研究報告，2007-MUS-072, pp. 25-28, 2007.

3) 酒向慎司, 宮島千代美, 徳田恵一, 北村正, "隠れマルコフモデルに基づいた歌声合成システム," 情報処理学会論文誌, Vol. 45, No. 3, pp. 719-727, 2004.

4) 吉田由紀，中嶌信弥，"歌声合成システム：CyberSingers ," 情報処理学会研究報告, 1998-SLP-025, pp. 35-40, 1998.

5) J. Sundberg, "The Science of Singing Voice," Northern Illinois University Press, 1987.

6) A. B. Meribeth, "Dynamics of the Singing Voice," Springer, 1997.

7) 大石康智, 後藤真孝, 伊藤克亘, 武田一哉, "スペクトル包絡と基本周波数の時間変化を利用した歌声と朗読音声の識別," 情報処理学会論文誌, Vol. 47, No. 6, pp. 1822-1830, 2006.

8) 辻直也，赤木正人, "歌声らしさの要因とそれに関連する音響特徴料の検討" 日本音響学会聴覚研究会資料, H-2004-8, 2004.

9) T. Saitou, M. Unoki and M. Akagi, "Analysis of acoustic features affecting "singing-nes" and its application to singing voice synthesis from speaking voice," Proc. ICSLP2004, Vol. III, pp. 1929-1932, 2004.

10) H. Kawahara, I. Masuda-Katsuse, A. de Cheveigne, "Restructuring speech representations using a pitch adaptive time-frequency smoothing and an instantaneous- frequency based on F0 extraction: Possible role of a repetitive structure in sounds," Speech Commun., Vol. 27, pp. 187-207, 1999.

11) T. Saitou, M. Unoki and M. Akagi, "Development of an F0 control model based on F0 dynamic characteristics for singing-voice synthesis," Speech Commun., Vol. 46, pp. 405-417, 2005.

12) H. B. Rothman, A. A. Arroyo, "Acoustic variability in vibrato and its perceptual significance," J. Voice, Vol.1, no.2, pp.123-141, 1987.

13) 齋藤毅. 鵜木祐史, 赤木正人, 榊原健一, "歌声の基本周波数変化に含まれるオーバーシュートの知覚への影響に関する検討，" 日本音響学会聴覚研究会資料, H-2006-109, 2006.

14) 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，" 日本音響学会聴覚研究会資料, TL2005-10, 2005.

15) M. Akagi and H. Kitakaze. "Perception of synthesized singing-voices with fine-fluctuations in their fundamental frequency fluctuations," Proc. ICSLP2000, Vol. 3, pp. 458-461, 2000.

16) 中山 一朗, "日本語を歌・唄・唄う," 日本音響学会誌, 59 巻, 11 号, pp. 688-693, 2003.

17) W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, "Numerical Recipes in C," Cambridge University Press, Cambridge, 1988.

18) 天坂格郎, 長沢伸也, 官能評価の基礎と応用：自動車における感性エンジニアリングのために，日本規格協会, 2000.

19) J. Sundberg, "Articulatory Interpretation of the 'Singing Formant'," J. Acoust. Soc. Am, Vol. 55, pp. 838-844, 1974.

20) I. Nakayama, "Comparative studies on vacal expression in Japanese traditional and western classical-style singing, using a common verse," Proc. ICA2004, Mo 4. Cl.1, pp. 1295-1296, 2004.

21) 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," 日本音響学会音声研究会資料, SP-89-147, 1990.

22) S. Wang, "Singer's high formant associated with different larynx position in style of singing," Journal of Acoustic Society Jpn. (E) 7, pp. 303-314, 1986.

23) K. Honda, T. Kitamura, H. Takemoto, M. Fujita, P. Mokhtari, "Resonance characteristics of hypopharyngeal cavities," Proc. ICVPB2004, 2004

24) J. Sundberg, "Singing and timbre," Music room acoustic, Stockholm: Royal Swedish Academy of Music Publications, Vol. 17, pp. 57-81, 1977.

25) Y. Horii, "Acoustic analysis of vocal vibrato: a theoretical intepretation of data," J. Voice 3, pp. 36-43, 1989.