

## 確率的手法による歌唱曲の自動作曲

嵯峨山茂樹\*・酒向 慎司†・堀 玄‡・深山 覚\*

## 1. はじめに

人間が歌の作曲をするように、コンピュータにそれをさせられれば、自作の詩を曲にしたり、それを恋人の誕生日に贈り物にしたり、脚本に作曲してミュージカルを作ったり、大切な事項を歌にして記憶したり、社歌を作曲したり、著作権の問題のないオリジナル曲をビデオやホームページで挿入したり、さまざまな利用が考えられよう。さらに、自分で作曲するユーザへは、作曲のヒント提供にも役立つ。

実は、コンピュータを作曲に利用する試みは早くからなされているが、多くは「アルゴリズム作曲」とよばれる現代作曲が目的であった。一方、一般のユーザが歌詞を与えて、普通に歌え、感動を与えられる曲を自動作曲する試みはまだ多くない。人間の高度な知能活動を機械にさせる情報処理分野のチャレンジといえるだろう。

本稿では、歌詞に基づく歌唱曲作曲をモデル化し、これに基づき任意の日本語テキストを歌詞とする歌唱曲を自動作曲する手法について解説し、著者らが開発した自動作曲システム *Orpheus* を紹介する。基本原理は、その旧バージョンの解説[1]は共通しているので、本稿 2., 3. の詳細はその文献を参照されたい。

## 2. 歌唱曲自動作曲のアルゴリズム

## 2.1 自動作曲の探索問題としての定式化

旋律にはどれくらい種類がありえるのか試算しよう。たとえば、「我輩は猫である」という歌詞は仮名では10文字であるから、4分の4拍子の2小節に10個の音符を配置すると、まず音高として仮名それぞれに、たとえばシドレミファソラシドレの10種類の音高のどれかを当てはめる組合せは $10^{10} = 100$ 億通りであり、それらを16分音符単位の32個の時点に配置する ${}_{32}C_{10} > 6000$ 万通りを考えると、60京種類以上ということになる。実際には歌詞はもっと長いうえに、さらに#やbを含めると膨大な組合せになることがわかる。自動作曲とは、この中から、ユーザの意向を反映して、音楽的であると感じさ

\* 東京大学 大学院 情報理工学系研究科

† 名古屋工業大学 大学院 工学研究科

‡ 亜細亜大学 / 理化学研究所

**Key Words:** automatic music composition, probabilistic models, prosody, singing voice synthesis, automatic lyrics generation.



第1図 イントネーションと旋律の関係：この例は「箸の端に橋があった。端の橋に箸があった。」と聞こえるのではないか？

せる確率を最大にする音符系列を見つける探索問題として捉えることができる。では、どうすれば音楽的であると感ぜられるのか。それには作曲理論が役立つ。

## 2.2 歌詞からの制約：抑揚

日本語はピッチアクセントの言語である。関東方言（いわゆる標準語）では、一般に  $m$  モーラの単語あるいはアクセント句には0型から  $m$  型までの  $m+1$  種類のアクセント型があり得る。0型はピッチの下降のないものである。ほかは、ピッチの下降の直前のモーラ（アクセント核）が何番目であるかにより、 $m$  型と称する。

たとえば「端」「箸」「橋」はいずれも2モーラ語で発音は「はし」であるが、アクセント型はそれぞれ0型、1型、2型である。「端」と「橋」の相違がわかりにくいかもしれないが、それぞれに助詞「を」をつけてアクセントを視覚的に表すと、 $\overline{\text{ハ}}|\underline{\text{シ}}\text{ヲ}$ ,  $\overline{\text{ハ}}|\underline{\text{シ}}\text{ヲ}$ ,  $\overline{\text{ハ}}|\underline{\text{シ}}|\underline{\text{ヲ}}$  となる。

たとえば「箸の端に橋があった」というテキストは、

$\overline{\text{ハ}}|\underline{\text{シ}}\text{ノ}$   $\overline{\text{ハ}}|\underline{\text{シ}}\text{ニ}$   $\overline{\text{ハ}}|\underline{\text{シ}}|\underline{\text{ガ}}$   $\overline{\text{ア}}\text{ツタ}$

とアクセント句に分割され、イントネーションがつけられる。第1図に示すような曲があるとすれば、イントネーションと旋律が一致していることが感ぜられるだろう。

日本語歌唱曲の作曲では、歌詞のイントネーションと旋律が一致することが要求されてきた[2]。山田耕筰らの作曲スタイルとしてもよく知られている。

旋律設計は、韻律の上下動を守りながら、曲想に応じてどの音域を使うか、どの程度の跳躍をするか、などを考慮して進められる。この際、音程だけでなく音価も同時に考えられる。強弱アクセント言語を歌詞とする場合と比べて、歌詞が日本語の場合はリズムの制約は弱いが、歌詞のモーラ数により音符数は決まるので、器楽曲に比べればリズム設計の自由度はかなり小さくなる。こうした制約の中で、ユーザが望む曲想に合わせて静かなリズム、躍動的なリズム、などを決定できる。

### 2.3 和声構造からの制約, 非和声音, 禁則

単旋律を作るならばまだしも, 伴奏を伴った曲を作るうえでは, 旋律の背後の和声進行が重要である. 和音と, その連鎖としての和声進行に関する理論は「和声学」とよばれ, 西洋音楽の規範であり, 作曲の基本原理である. 西洋音楽において多声音楽が対位法から出発し発達して, 同時に鳴る音の成す響きに注意が移り, 17世紀前後に次第に確立したのが和声学である.

本稿では和声学 [2,3] を詳しく述べることはできないが, 要点を述べれば, 和音は根音の上に三度音程で積み重ねた三つの音で構成される三和音や四つの音で構成される七の和音を基本とする. 根音が同じでも積み重ねる三度音程の長短によりいろいろな印象の和音の響きが生じる. 最も重要なのは, これらの多種の和音を順次連結することにより生じる和声進行である.

和声内音だけで旋律を作れば和声とも合致するし, 基本的な制約はない. しかし, それでは滑らかな動きの旋律など, 豊かな旋律表現には材料不足である. ある和音が響いているとき, その和音の構成音以外の音が使われるとき, その音を非和声音といい, 和声学で使用が許されている非和声音には, 経過音, 補助音 (刺繍音), 倚音, 繋留音, 先行音, 逸音などがあり, これらを適切に使えば, いわゆる旋律らしい音符列を生成できる.

和声学には声部進行の禁則がある. 歌唱と伴奏など, どの声部間でも連続 (平行) 同度, 五度, 八度が禁止され, おもに外声部間で並達 (陰伏) 同度, 五度, 八度が禁止される. 最近のポピュラー曲ではこれらの禁則を守らず, 聴き苦しい曲が散見されるが, せめて外声部の禁則違反は目立つので守りたい. その場合は, 伴奏の低音進行に配慮してそれとの間で声部進行の禁則を犯さないようにするという制約が生まれる.

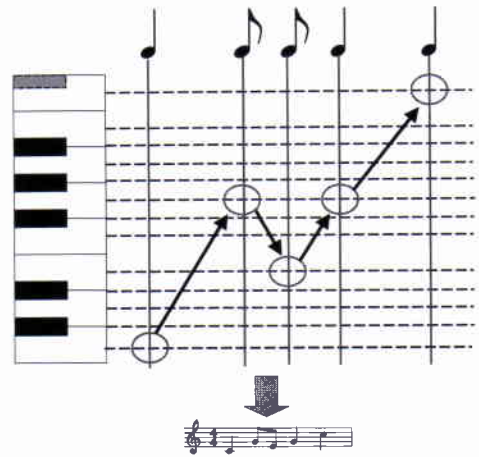
そのうえで, 和音に分散和音形などの音型を与えることで, 伴奏に和声, リズム, 曲想の3点で旋律を補助する役割をもたせられる. とくに和声の面で, 伴奏は旋律だけでは表現しきれない和音の特色や豊かな響きを作ることができる.

### 2.4 動的計画法を用いた旋律設計

以上の議論から, 曲想を基にユーザが和声, リズム, 伴奏音型を独立に選択し, そこから旋律および伴奏を設計するという順序で歌唱曲作曲のプロセスが考えられる.

旋律を設計する声域と, 旋律のリズムが与えられたとき, 旋律が取りうる音程の候補と, その候補の中から使用する音程を選択するタイミングが決定される. このとき, 旋律は第2図に示されるように, 音程とリズムからなる格子点をリズムの進行にしたがって遷移する経路と捉えることができる.

以上に述べたさまざまな音楽的制約を尊重し, 歌詞の韻律を最大限活かして作曲を行うならば, それが良い曲と感じられる確率が高い, とすると, その確率を最大化



第2図 経路としての旋律決定の例. 図上部のようにリズムが与えられると, それと音高から形成される格子点上を図中のように遷移する経路は, 図下部に示した旋律である.

する経路を探索する問題を解けばよい. ユーザが満足する確率が最も高い旋律はどのような経路か, とたとえられよう. この格子点の確率最大経路探索問題は, 動的計画法により, 歌詞の長さ按比例する計算量で効率的に解くことができる.

## 3. 自動作曲システム Orpheus

以上に述べた作曲の原理に基づき, 自動作曲システム Orpheus を構築し [5], その後も継続的にシステムの改良を重ねている. 処理の流れを第3図に示す. 以下に処理手順を示す.

### (1) ユーザによる入力と選択

ユーザは, 漢字かな混じりの日本語テキストを入力するとともに, 自分が望む曲想から, テンポ, リズム (木構造), 和声進行, 合成音声, 伴奏音型, 伴奏音色, ドラムス, などの選択肢から選択する. これらの作曲条件のセットが多数用意されていて, ユーザには容易にスタイル選択ができる.

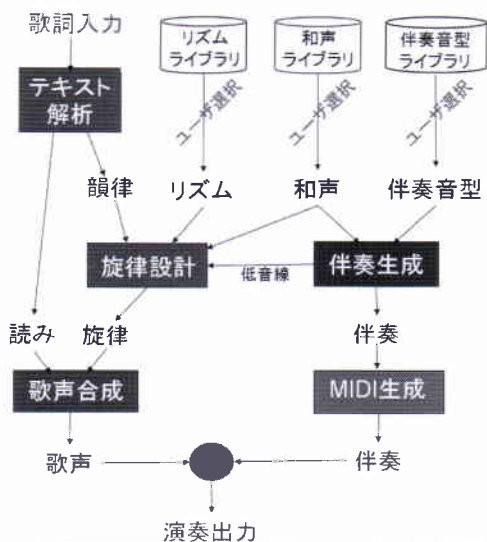
### (2) テキスト解析

テキスト解析部では, 入力された日本語テキストに対し, テキスト朗読時の読み, 韻律を解析する.

解析ツールとして, Galatea Project [4] で開発されたテキスト音声合成システム “GalateaTalk” における日本語テキストを解析して韻律情報を抽出する部分を利用した. 解析されたテキストは, 2小節を単位とする節に分割される.

### (3) リズムの決定

以上の処理により, 各節の音符数が決定するので, これを基にユーザが選択した木構造のリズム (リズム木) から音符数が合致するリズムパターンを抽出する. リズム木は, 統一感のあるリズムを生成するための, われわれの「リズム木仮説」に基づくデータ構造であり, 現在



第3図 Orpheusのシステム構成

はあらかじめ人手で作成しているが、音楽データからの自動生成も可能であり、今後の研究テーマである。

(4) 伴奏生成

伴奏は、和声とともに与えられる伴奏構成音と、ライブラリからユーザが選択した伴奏音型を用いて生成される。生成された伴奏は、テキストベースで楽譜記述言語として出力され、楽譜出力、MIDI出力を行う。

(5) 旋律設計

すでに述べた手法により、動的計画法による確率最大経路探索問題により旋律を決定する。

(6) 統合出力と音響出力

本システムで生成された旋律および伴奏は、楽譜出力とMIDI出力される。また、後述する歌声合成サブシステムにより旋律の歌声を合成し、これと伴奏を信号領域で加算することで伴奏付きの歌声による自動演奏出力を行う。

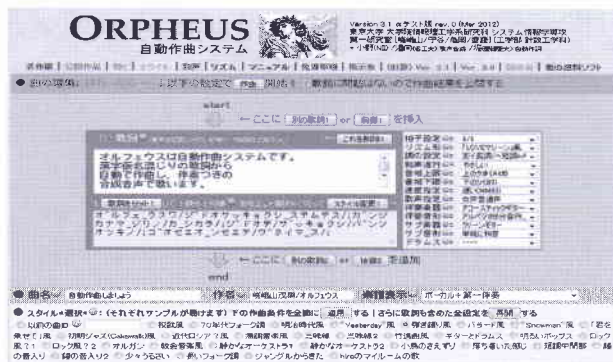
このようなシステムを、第4図に示すようなwebインタフェースを通してユーザが利用できるように実現した。前奏・間奏・後奏も含む多楽節の曲を容易に作れる。短い自動作曲の出力楽譜の例を第5図に示す。

4. 確率的言語モデルに基づく自動作詞

4.1 作詞の自動化

作詞は素人にとってはそれほど簡単な作業ではない。われわれは、ユーザが入力した少数のキーワードをもとに歌詞を自動的に生成するシステムを、確率的言語モデルに基づき実現することを考えた。

この分野の先行研究としては、MITのR. Kurzweilによるマルコフモデルに基づき詩を生成するシステムRKCP (R. Kurzweil's Cybernetic Poet)[7]がよく知られている。また、日本語の歌詞を作成するソフトウェアの事例のPlayStation2用ゲームソフト「くまうた」は、クマがランダムに作った歌詞をプレイヤーが修正してい



第4図 自動作曲システム Orpheusの web インタフェース



第5図 自動作曲システム Orpheusの出力楽譜例。入力歌詞は「オルフェウスは自動作曲システムです。漢字仮名混じりの歌詞から、自動で作曲し、合成音声で歌います。」

き、完成した曲をクマ（音声合成システム）が歌うものである。

われわれは、確率的言語モデルの一つである bigram モデルに基づき歌詞を自動的に生成するサブシステムを Orpheus に追加したので、以下に説明する。原理は、bigram モデルの遷移確率に加え、離れた語彙の関連性をとらえるために共起確率を使い、動的計画法によりユーザが入力したキーワードを含む歌詞を生成するものである。

4.2 Bigram モデルの推定

単語の出現確率が直前の単語だけに依存すると仮定するモデルを bigram モデルとよぶ。大量の歌詞コーパスが得られたとすると、bigram モデル確率の推定のために、まず歌詞コーパスの各行を形態素解析する。形態素解析とは、与えられた文を形態素（意味をなす最小の単位）に分割する処理であり、これを基に相対頻度として形態素間の遷移確率を推定するのだが、たまたま歌詞コーパスに現れなかった形態素間の遷移は確率がゼロになるという「ゼロ頻度問題」を回避するために、遷移確率の修正法である「平滑化（スムージング）」として、広

く使われているグッド・チューリング法を用いている。

### 4.3 共起確率

前節で求めた遷移確率だけを使って歌詞の生成を行うと、歌詞コーパスのもつ情報のうち二つ以上離れた形態素間の関連性に関する情報を捨てていることになる。しかし、 $N \geq 3$ の連鎖統計として  $N$ -gram モデルを使うことは、データ量の点からも、歌詞生成時の計算量の点からも、現実的でない。そこで本研究では、bigram モデルに基づきながら離れた形態素間の関連性も取り入れるために、共起確率の考え方を取り入れて、歌詞を生成する際に、ユーザが入力したキーワードとの共起確率が高い形態素は、比例して高い確率で生起するようにしている。共起確率を求めるために、前節で形態素解析した歌詞コーパスから、各形態素のキーワードとの行単位の共起回数（文始端 BOS から文終端 EOS の区間で共起している回数）を数え上げる。これをキーワードを含む行の数で割り、共起確率を求めている。

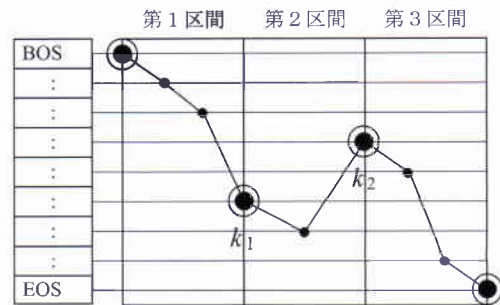
### 4.4 動的計画法による歌詞の生成

歌詞コーパスを形態素解析して得られた形態素リストから適切なものを選択し配列して新たに歌詞を生成する作業は、形態素リスト上の経路探索問題としてとらえることができる。ここではユーザが入力したキーワードが入力順に現れる歌詞を生成したいので、BOS から始まりキーワードを順に通過し EOS で終わる経路を探索することになる。前節までに求めた遷移確率と共起確率の値をもとに、このような経路で確率が最大となるものを求めることにより、歌詞コーパスのもつ形態素の配列の特徴を保ちつつ、所定のキーワードを含む歌詞を生成することができる。

この経路探索問題は、音声認識で用いられる 2 段 DP [8] を一部変更した次のような手順で効率的に解くことができる（第 6 図）。まず、ユーザが入力したキーワード  $k_1, \dots, k_n$  に便宜的に  $k_0 = \text{BOS}$  と  $k_{n+1} = \text{EOS}$  を付加し、 $k_{i-1}$  と  $k_i$  をつなぐ各経路（第  $i$  区間とよぶ）を動的計画法により探索する。ここでは経路上の形態素数にある程度の幅をもたせ、各区間について経路の候補を複数求める。また計算時間の短縮のためにビームサーチを行う。つぎに、各区間について求めた候補を総当たりで組み合わせて 1 行の歌詞とし、ユーザが指定した文字数からのずれが一定の値より小さいものを最終的な候補とする。最後に、最終的な候補の中で確率が大きいものから順に 10 個を出力とする。

### 4.5 自動作詞システムの実装

前節で述べた歌詞生成の原理に基づき、自動作詞システムを実装した。ユーザは、5 個までのキーワード、歌詞の 1 行の文字数、歌詞の行数を指定して作詞を行うことができる。歌詞の行数が複数行になる場合は、キーワードは適当に各行に分散される。また、4.2 と 4.3 で述べた遷移確率と共起確率を求める作業を「ポップス」



第 6 図 経路探索による歌詞生成

と「演歌」の歌詞コーパスに対して行っており、ユーザはいずれかのジャンルを選択して作詞を行うことができる。ユーザが入力したキーワードは形態素解析され、名詞、動詞、形容詞だけが残される。また、動詞、形容詞は原形に変換され、原形がこれと一致する形態素がキーワードとして使われる。作成した歌詞は、自動作曲システム Orpheus の歌詞入力欄に転送できる。

以下は、「ポップス」と「演歌」のそれぞれのジャンルを選択し、キーワードとして「夢」と「恋」の二つを指定し、1 行の文字数を約 22 文字として 1 行の歌詞を生成した例である。

「ポップス」 キーワード=夢, 恋 文字数=約 22 文字  
 小さな夢の恋堕ちる時。  
 夢を向いてるオレンジな恋。  
 小さな夢の恋堕ちる時だろう。  
 小さな夢の涙恋ならばいい。

「演歌」 キーワード=夢, 恋 文字数=約 22 文字  
 二度とかえらぬ夢を忘れられない恋に。  
 夢を忘れられないしよどない恋だった。  
 夢が悪いわけじゃないしよどない恋に。  
 ああ夢を忘れられない恋のように。

## 5. 確率モデルに基づく歌声合成

### 5.1 自動作曲システムのための歌声合成

最後に、歌詞とメロディから歌声の音声波形を合成する技術について述べる。Orpheus から歌唱曲が生成されると、メロディと各音符に対応する歌詞の読みが定まる。したがって、日本語の音韻を任意の長さ、音階で発声できる音声合成システムがあれば、ボーカルパートの音響データを得ることは可能である。しかしこれだけでは、満足のいく歌声にはならないであろう。

たとえば、ピアノ演奏を MIDI の打ち込みで作成する場合には、メロディに沿ってノートオン・オフを並べるだけでは、単調で機械的な演奏になるだけである。実際には、強弱の変化や発音のタイミングや長さを適切に調整することで、実際の演奏のような表情のある演奏を作ることができる。音楽情報処理の分野では、それを自動

化するための演奏表情付与の研究も盛んに行われている。

人間の歌声の場合でも、メロディに沿って忠実にピッチを変化させることはむしろ不可能であり、人間の歌声らしい表情が存在し、歌い手の個人性や歌唱スタイルが反映されたものである。そこで、任意の歌唱曲から表情豊かな歌声を自動合成できる技術が必要である。一方で、音楽制作やアミューズメントの観点では、歌声の装飾や加工にユーザが自由に介入できる余地も重要な要素であろう。

以上をまとめると、Orpheusにふさわしい歌声合成システムの要件として、個性ある人間の歌声を自動合成できるとともに、ユーザの要求に合わせて歌声を加工することができる柔軟な制御性が重要である。このような指針のもとで、声質や歌唱法などの「歌い手の特徴」を歌声データから自動学習し、それらを再現することが可能な隠れマルコフモデル (Hidden Markov Models; HMMs) に基づいた歌声合成法について述べる [9]。

## 5.2 HMM に基づいた歌声合成

HMM 歌声合成システムは、第7図に示す概略のように、おもに学習部と合成部の二つのパートから構成される。

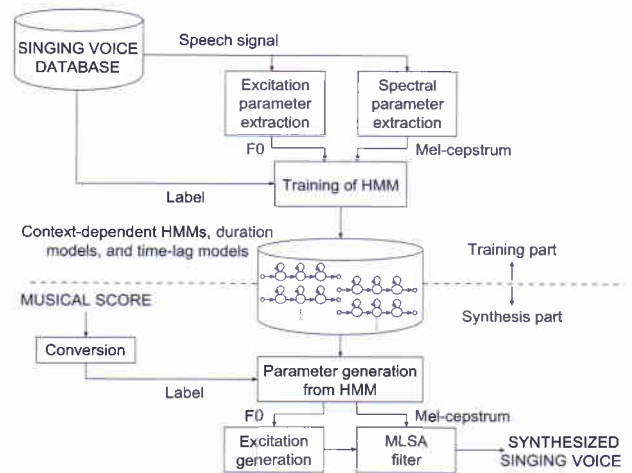
元来はテキスト音声合成の技術 [11] を、歌声の合成を主眼に拡張したものであり、歌声合成のモデルパラメータを、一定量の歌声データ学習によって得ることが大きな特徴である。

学習部は、歌声データベースから歌声合成用のモデルである HMM を学習するもので、Orpheus の処理とは別のオフラインパートである。この歌声データベースには、一定量の曲を人が実際に歌った波形データと、それぞれの曲の楽譜情報が含まれる。伴奏などを伴わないボーカルパートだけの歌声データで、かつ一定量の規模となると、このようなデータベースは一般的には流通していない。Orpheus に組み込まれている男女の歌声モデルは、研究用に独自に収録した歌声データベースから学習されたもので、童謡や唱歌を対象とした 60 曲程度の歌声が対象となっている。なお、歌声データベースの作成方法については、文献 [10] などで示されている。

合成部では、学習パートで得られた HMM を利用し、合成したい歌唱曲から歌声の音声波形を生成する。この入力情報は、基本的には歌詞の読みとメロディだけであり、ユーザが指定することのできる歌声合成用のパラメータを伴って Orpheus の内部処理として組み込まれている。

## 5.3 制御性を備えた歌声のモデル化

人間の発声を模したモデルとして、声帯の振動による音源 (ソース) の生成と、声道形状 (フィルタ) の変形による調音制御に分離したソース・フィルタモデルがある。このようなモデルにおいて、音声信号から音源と声道特性を分析抽出し、そこからもとの信号に再合成する



第7図 HMM に基づく歌声合成システムの概要

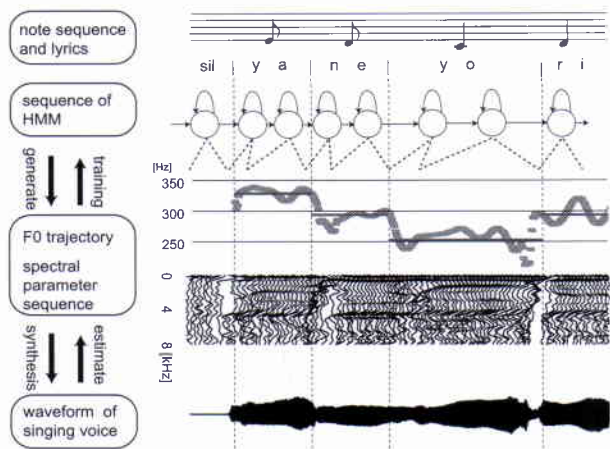
分析合成技術は、現在の音声信号処理の基盤技術となっている。

この仕組みを歌声に照らし合わせると、メロディに沿って声帯振動を制御し、同時に歌詞に沿って調音制御を行うことに相当する。HMM 歌声合成システムでは、あるメロディと歌詞の歌声を合成するために、声帯の振動や声道特性に相当するパラメータ列の生成過程を、統計モデルで記述したものと考えることができる (第8図)。具体的には、歌声の基本周波数 ( $F_0$ ) と、スペクトルの概形 (ケプストラム係数) の系列を歌声データから推定し、その時間変化を言語音の最小単位である音素ごとに学習する。歌声を合成する際には、歌詞とメロディによって定まる音高と音素列に沿ってこれらのモデルを接続することで、与えられたメロディと歌詞に応じた歌声のパラメータ列を生成し、歌声の波形データが出力される。

このようにして、音声の合成過程をパラメトリックなモデルで記述することによる利点として、どのようなメロディや歌詞でも自由に歌わせることができるほか、ピブラートのような  $F_0$  の揺らぎを付与したり、声道特性を変化させることで声質を調整するような操作が容易に実現できる。Orpheus の合成システムにおいても、いくつかのパラメータによってユーザが歌声を調整することができる。

## 5.4 個人性を備えた歌声のモデル化

音素は言語音を分類する基本要素であるが、そのスペクトルや  $F_0$  の形状はさまざまな要因によって変動を伴う。たとえば、音素のつながりによる音韻の変形、母音の無声化など音響特徴が変化することが知られている。歌声について考えてみると、たとえば異なる音高で /a/ を発声した場合に、ピッチ以外の声道成分やその時間変化に影響がないわけではなく、低音域と高音域ではスペクトルの概形が異なる場合が多い。また、音符の長さによっても少なからず影響を受けると考えられ、これらの特徴は、歌い手や歌唱スタイルの違いによっても異なっ



第8図 HMMから歌声が合成される過程の概要

た傾向をもつと考えられる。したがって、人間らしい個性を伴った歌声の合成には、このような違いを適切に再現することが重要となる。

HMM 歌声合成では、歌声の音響的な変動要因を「コンテキスト」とよび、コンテキストに依存した詳細なモデル化を行っている。つまり、同じ/a/という発声でも、ピッチや音長が異なる場合には、別の歌声として扱うことで、これらの変動に対応することができる。詳細に分類することにより、その組合せが膨大な数となるため、事前に学習データを確保できない問題が生じる。実際には区別する必要がない（音響的な違いが生じない）場合も存在し、分類する必要性の高さを考慮して効率的にクラスタリングするアルゴリズムが導入されており、現実的な規模の学習データから、このような詳細な分類のモデルを学習することができる。

## 6. おわりに

本稿では、歌詞の韻律を旋律設計に反映した歌唱曲の自動作曲、それに付随する自動作詞、歌声合成を、すべて確率最大の最適経路問題として定式化した手法を紹介した。この手法をもとに任意の日本語歌詞を入力とし、ユーザに選択された和声、リズム、伴奏音型などを制約条件として歌唱曲の自動作曲を行い、歌声と伴奏と自動演奏出力を行うシステム Orpheus を構築した。このシステムは web ベースの試行サービスとしてたびたびバージョンアップされ [6]、一般ユーザに供されており、作成された曲は累積で 10 万曲を超える。最新版は以下の URL

<http://www.orpheus-music.org/v3/>

で試行できる。

(2012年2月22日受付)

## 参考文献

- [1] 嵯峨山, 中妻, 深山, 酒向, 西本: 日本語歌詞からの自動作曲; オペレーションズ・リサーチ: 経営の科学, Vol.

54, No. 9, pp. 546-553 (2009)

- [2] 長谷川: 作曲法教程 上巻, pp. 68-89, 音楽之友社 (1950)
- [3] 島岡: 和声と楽式のアナリゼ, pp. 8-9, pp. 18-20, pp. 37-39, 音楽之友社 (1964)
- [4] S. Kawamoto, et al.: Galatea: open-source software for developing anthropomorphic spoken dialog agents; *Life-Like Characters, Tools, Affective Functions and Applications* (H. Prendinger et al. (Eds.)), Springer, pp. 187-212 (2004)
- [5] 中妻, 酒向, 小野, 嵯峨山: 歌詞の韻律を用いた自動作曲; 日本音響学会春季研究発表会講演論文集, pp. 739-740 (2007)
- [6] 深山, 中妻, 米林, 酒向, 西本, 小野, 嵯峨山: Orpheus 歌詞の韻律に基づいた自動作曲システム; 情報処理学会研究報告, 2008-MUS-76, pp. 179-184 (2008)
- [7] R. Kurzweil: *The Age of Intelligent Machines*, MIT Press (1992)
- [8] 迫江, 千葉: 動的計画法を利用した音声の時間正規化に基づく連続単語認識; 日本音響学会誌, 27 巻, 9 号, pp. 483-490 (1971)
- [9] 酒向, 宮島, 徳田, 北村: 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)
- [10] 酒向, 北村: HMM 歌唱合成のための歌唱データベースの構築; 日本音響学会 2010 年秋季研究発表会講演論文集, 3-P-30, pp. 399-400 (2010)
- [11] 吉村, 徳田, 小林, 北村: HMM に基づく音声合成におけるスペクトル・ピッチ・継続長の同時モデル化; 信学論 (D-II), Vol. J83-D-II, No. 11, pp. 2099-2107 (2000)

## 著者略歴

嵯峨山茂樹



1972年東京大学工学部計数工学科卒業, 1974年同大学院工学系研究科計数工学専攻修士課程修了。同年日本電信電話公社(現NTT)入社, 武蔵野電気通信研究所にて音声情報処理の研究に従事。1990-1993年 ATR 自動翻訳電話研究所音声情報処理研究室長として自動翻訳電話の開発に従事。1998年北陸先端科学技術大学院大学情報科学研究科教授。2000年東京大学大学院情報理工学系研究科システム情報学専攻教授, 現在に至る。音楽・音声・音響・文字の信号処理・情報処理の教育研究に従事。博士(工学)。発明協会発明賞, 科学技術庁長官賞, 電子情報通信学会論文賞, 情報処理学会論文賞などを受賞。電子情報通信学会フェロー, 日本音響学会, SICE, IEEE, ISCA の各会員。現 IEEE Signal Processing Society Japan Chapter Chair。

酒向 慎司



1999年名古屋工業大学工学部知能情報システム学科卒業，2004年同大学院博士後期課程修了．同年4月東京大学大学院情報理工学系研究科特任助手を経て，2007年名古屋工業大学大学院工学研究科情報工学専攻助教となり，現在に至る．音楽情報処理，手話認識などの研究に従事．博士(工学)．2011年電子情報通信学会ヒューマンコミュニケーション賞受賞．電子情報通信学会，日本音響学会，情報処理学会などの会員．

深山 覚



2008年東京大学理学部地球惑星物理学科卒業，2010年東京大学大学院情報理工学系研究科システム情報学専攻修士課程修了，現在同博士課程在籍．2011年日本学術振興会特別研究員(DC-2)に採用され，今に至る．自動作曲，自動演奏表情付けなど，音楽コンテンツを自動生成するための研究に従事．2008年情報処理学会音楽情報科学研究会「夏のシンポジウム」ベストプレゼンテーション賞，2009年情報処理学会山下記念研究賞受賞．IEEE，情報処理学会，日本音響学会，日本音楽認知学会，先端芸術音楽創作学会の各学生会員．

堀 げん



1991年東京大学工学部計数工学科卒業，1996年東京大学大学院工学系研究科情報工学専攻博士課程修了．(独)理化学研究所研究員，同副チームリーダーを経て，2009年亜細亜大学経営学部准教授となり，現在に至る．情報幾何学，信号処理，遺伝子情報処理などの研究に従事し，最近は音楽情報処理の研究にも着手．博士(工学)．IEEE，電子情報通信学会などの会員．