

数理解析研究所講究録 第 1743 巻 2011 年 1-5

数理解析研究所講究録 第 1743 巻 2011 年 1-5

片寄晴弘(関西学院大学)

Haruhiro Katayose (Kwansei Gakuin University)

1. はじめに

音楽は、私たち現代人にとってスポーツと並んで、最も身近なエンターテインメントである。その音楽の聞き方について、普段、我々は意識することはほとんどないが、技術革新の影響を大きく受けた結果として成立している。19 世紀末にレコードとラジオが相次いで発明され、20 世紀になって普及した。ラジオとレコードは、音楽の時間芸術としての制約、つまり、その場に居合わせなければ聞くことができないという制約を取払った。その結果、我々は、場所・時間を超えていつでもどこでも音楽を聞くことができるようになった。もちろん、ライブ音楽にはライブ音楽にしか味わえない良さがあり、その愉しみのスタイルは健在しているが、レコードやラジオ技術によって生まれた鑑賞のスタイルは当たり前のものであるとして受け止められている。

現在、メディア技術の発展、ブロードバンドインターネットの普及に支えられ、音楽コンテンツがデータとして流通するようになった。今や、ほぼ無尽蔵の音楽タイトルを聞くことができるようになりつつある。また、YouTube やニコニコ動画等の動画共有サイトサービスに連動する形で、今まで消費者だった一般ユーザによるコンテンツ制作が一挙に広まりまった。現代は、19 世紀末から 20 世紀の音楽の文化革命に匹敵、あるいは、それ以上の音楽の文化革命期と言える可能性がある。

このような時代背景のもと、CrestMuse プロジェクトでは、ヒトが音楽を聞いた経験、つまり、「音楽の事例」に着目して、能動的音楽鑑賞インタフェース、さらに、「事例」を操作して音楽を作るインタフェースの開発を進めてきた。本稿では、その概略を紹介する。

2. CrestMuse プロジェクトの取り組みの概略

CrestMuse プロジェクト(研究代表者:片寄晴弘)は、科学技術振興機構 CREST「デジタルメディア作品の制作を支援する基盤技術(研究総括:原島博)」領域で採択された音楽系の研究プロジェクト「時系列メディアのデザイン転写技術の開発」の略称であり、2005 年から 2010 年にかけて研究が実施された。

CrestMuse プロジェクトでは、ヒトが音楽を聞いた経験、つまり、「音楽の事例」に着目し、**音楽のデザイン転写**という新たなキーコンセプトを提唱し、音楽コンテンツの分析・生成・能動的鑑賞の基盤技術ならびにアプリケーションの開発に取り組んできた。ここでいう「デザイン転写」とは音楽コンテンツの単純なコピーをさすものではない。デザインをそのデザインに足らしめている特徴を抽象化して取り扱うことを対象としている。併せて、そのデザインのユーザ操作手段として **directability** というキーコンセプトを打ち出し、システムの開発を進めてきた[1]。

技術を縦軸とするなら音楽は横軸に相当する。音楽には、「聞く」「弾く」「作る」など音楽への異なったアクションが複数あり、また、それらをカバーするために、信号処理やシンボリック処理など異なった技術基盤を扱う必要がある。CrestMuse プロジェクトでは、「縦軸」と「横軸」からなる音楽情報処理技術をほぼ網羅する形で研究を進めてきた。ここでは、経時的に流れを示すという目的にたち、音楽デザイン支援技術と楽音分析処理技術の2系等にわけ、年表形式で成果の概要を紹介する。

2.1 音楽デザイン支援技術

図1に音楽デザイン支援技術視点で整理した成果を示す。この図において、上部にUGC(=User Generated Content)とCGM(=Consumer Generated media), 下部にメディアデバイスの普及状況を描き、それに対応させる形で CrestMuse プロジェクトの成果を整理している。

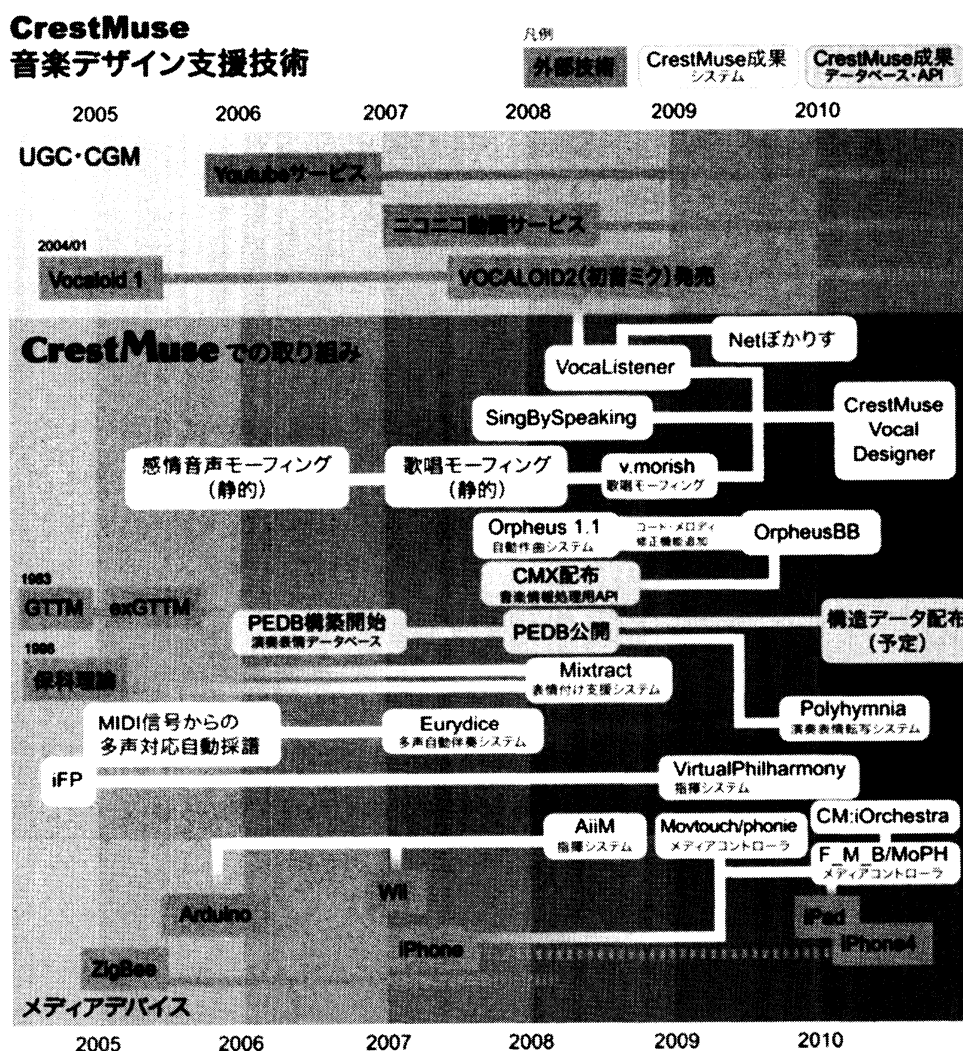


図1. CrestMuse プロジェクトの成果(音楽デザイン支援技術)

コンテンツの分析・生成に関する具体的な成果としては、a) CrestMuse Vocal Designer: 話声を歌

声に自動変換する SingBySpeaking, ユーザ歌唱を真似て歌声合成する VocaListener, 歌声を混ぜるインタフェース v.morish の三つのシステムの統合した歌声合成・変換システム, b) 歌唱スタイルレゾネータ: 一般歌唱を島歌風、ロックボーカリスト風に変換するシステム, c) OrpheusBB: 歌詞の韻律に基づいてメロディを作り, さらにユーザのメロディやコードなどの事後 editing に対して音楽的な整合性を保持する自動作曲システム, d) Mixtract: フレーズ構造、頂点音の提示支援に基づく演奏の表情付けシステム, e) VirtualPhilharmony: オーケストラが持つ音楽ヒューリスティックスを再現した指揮システム, f) CM:iOrchestra: Phone や iPad を用いた一連のメディアパフォーマンスインタフェース, などがあげられる。

2.2 楽音分析処理技術

CrestMuse プロジェクトの成果は, 前節で紹介したようにアプリケーションに注目が集まることが多いが, その基礎を支える多重音分析研究, 歌唱分析合成技術などの基礎信号処理領域の研究にも大きな進展があった. 図 2 に楽音分析処理技術視点で整理した成果を示す. この図に示すように, CrestMuse プロジェクトにおける楽音分析処理技術は, 歌唱(単声)を対象とした STRAIGHT 系の技術と, 多重音分析系の技術に大別される。

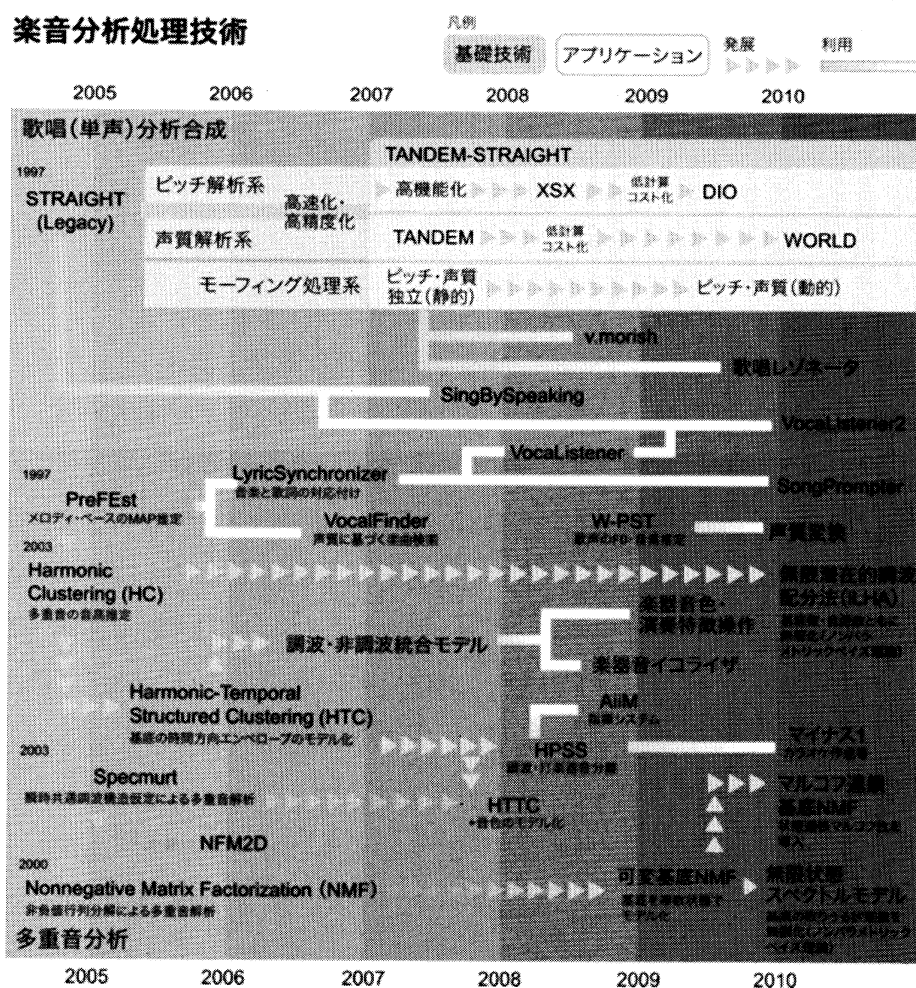


図 2. CrestMuse Project の成果(楽音分析処理技術視点)

STRAIGHT とは、Vocoder ベースの高品位の音声の分析・合成手法である。CrestMuse プロジェクトの研究期間中、よりシンプルな方式となる TAMDEM-STRAIGHT[2]の発明がなされ、あわせて、モーフィング処理に適した処理系の API が作られた。上述の CrestMuse Vocal Designer、歌唱スタイル(島歌風)レゾネータは、STRAIGHT を基盤技術として実装されている。

多重音分析系技術とは多声音楽の音響ストリームから各楽音を分離、操作する技術である。時間周波数平面に拡散した観測エネルギーパターンから、楽音を構成するパラメータ、および、どのタイミングにどの音が発音されたか、の双方がわからない状況で、同時推定することを目的とするこの課題はいわゆる不良設定問題である。ここでは、倍音によって構成される調波構造音と非調波構造音の組み合わせとして楽音をとらえ、さらに、ビブラートレベルの時間揺らぎが存在するか否かを数理的に記述して、楽音を分離・分析する技術の開発を行った。その代表的な成果として、亀岡らの g) HTC[3]があげられる。これらの多重音分析系技術研究を応用し、h) 楽器音イコライザや、i) Euterpe: マイナス1(カラオケ)作成アプリケーション、が作られた。この他にも楽音分析処理技術を基盤として、数多くの能動的音楽鑑賞システムの提案がなされている[4]。

3 展望:まとめにかえて

冒頭でも述べたように、2005 年から 2010 年にかけて音楽を取り巻く技術環境は大きな変遷を遂げた。文化・社会に関する事項として、この5年間で、1)UGC, CGM と 2)新世代メディアデバイスが、一般レベルで普及した。この間、ブロードバンドインターネットの普及とともに、YouTube やニコニコ動画等の動画共有サービスが開始され、UGC, CGM を加速させる原動力となった。動画共有サイトにおけるボーカル音源「初音ミク(vocaloid)」に端を発する歌唱デザインへの盛り上がりはインターネット文化における社会現象の一つとなっている。また、このことにより研究領域としての歌唱分析合成技術[5]や、著作物の取扱いに社会一般の目が向いたことも特筆に値する。

UGC, CGM の普及とともに、もう一つ忘れてはならないのは、新世代メディアデバイスの普及である。2006 年に任天堂から Wii が発売され、それ以降、ゲーム機器における PUI(Perceptual User Interface)の開発は時代の潮流となっている。また、2007 年の iPhone に始まり、2010 年度にはスマートフォンが大きく普及した。スマートフォンには、通信機能と AudioVisual 機能はもちろんのこと、加速度センサやカメラ、マルチタッチなどのセンシング機能が搭載されている。これらの機能は研究領域ではさほど驚くべきものではないが、低価格が進み普及レベルに達したというインパクトは思いのほか大きい。これらのデバイスを利用した新たな表現文化や教育メソッドが生まれてくる可能性は極めて高い。

最後に、もう一点取り上げたいのは、機械学習の新しい潮流を取り入れた音楽情報学の可能性である。楽音分析処理技術の取り組みにおいて、人間(システム開発者)の恣意的な制約や操作を

排し、できるだけ原理原則的な制約に基づいて処理を実現するという方向に向けて研究が進められてきた。その中で、ノンパラメトリックベイズモデルを応用した手法が最近提案され、「汎用性」確保という視点において有望視されている。ノンパラメトリックベイズモデルでは、生成系のモデルを見通しよくデザインするのにグラフィカルモデルを利用する。この方式に準拠することで、さまざまな音楽生成プロセスにおける共通要素を洗い出すことができる可能性が高い。現在、音楽情報処理領域におけるノンパラメトリックベイズモデルの応用は分析処理を対象としたものに限られているが、デザイン・生成系での応用を試行していきたい。

参考文献

- 片寄晴弘：音楽における自動処理とDirectability, オペレーションズ・リサーチ, Vol.54, No.9, pp.540–545 (2009).
- 2) Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T. and Banno, H.: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0 and aperiodicity estimation, *Proc. ICASSP*, pp.3933–3936 (2008).
- 3) Kameoka, H.: *Statistical Approach to Multipitch Analysis*, 東京大学大学院学位論文 (2007).
- 4) 後藤真孝：音楽音響信号理解に基づく能動的音楽鑑賞インタフェース, *IPSJ SIGMUS*, Vol.37, pp.59–67 (2007).
- 5) 後藤真孝, 齋藤 毅, 中野倫靖, 藤原弘将：歌声情報処理の最近の研究, 日本音響学会誌, Vol.64, No.10, pp.616–623 (2008).