

初音ミクとN次創作に関連した音楽情報 処理研究

VocaListenerとSongrium

Music information processing research related to Hatsune Miku and N-th Order Derivative
Creation

VocaListener and Songrium

後藤 真孝¹ | 中野 倫靖¹ | 濱崎 雅弘¹

GOTO Masataka¹; NAKANO Tomoyasu¹; HAMASAKI Masahiro¹

1 産業技術総合研究所 情報技術研究部門 (〒305-8568 茨城県つくば市梅園1-1-1 中央第2 産業技術総合研究所 情報技術研究部門) E-mail: m.goto[at]aist.go.jp

1 Information Technology Research Institute, National Institute of Advanced Industrial Science and Technology (AIST) (IT, AIST, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568)

原稿受理 (2013-12-05)

情報管理 56(11), 739-749, doi: 10.1241/johokanri.56.739 (<http://dx.doi.org/10.1241/johokanri.56.739>)

著者抄録

本稿では、初音ミクとN次創作に関連した音楽情報処理の研究事例として、歌声合成技術 VocaListener と音楽視聴支援サービス Songrium を紹介する。VocaListener は、初音ミク等の歌声合成ソフトウェアを用いて、録音された人間の歌声の事例からその歌い方（声の高さと声の大きさ）をまねて自然な歌声を合成する技術である。Songrium は、オリジナル楽曲とそれらのN次創作結果である派生作品といったさまざまな関係性を可視化できる音楽視聴支援サービスである。本稿ではさらに、より未来を見据えて自動創作と自動鑑賞の可能性を考察することで、音楽がもたらす感動の本質的な要因についても議論する。

キーワード

歌声合成, 初音ミク, CGM, N次創作, 音楽情報処理, 歌声情報処理

1. 初音ミクとN次創作

「人間の歌声でなければ聴く価値がない」という旧来の価値観が打破され、「合成された歌声がメインボーカルの楽曲を積極的に楽しむ文化」が世界で初

めて日本に誕生した。これは、ヤマハ株式会社の歌声合成技術「VOCALOID2」^{1)~3)}に基づいてクリプトン・フューチャー・メディア株式会社が歌声合成ソフトウェア「初音ミク」^{4)~6)}を2007年8月31日に発売したことによって引き起こされた。このソフトウェアを購入すれば、誰でも自分の楽曲を歌わせることができるため、アマチュアを中心とした多くのクリ

動画は、HTML版でご覧ください。

エーターによって、合成された歌声をメインボーカルとする楽曲が次々と発表された。

しかも、発売元が「初音ミク」という名前の仮想の歌手をイメージしたアニメ風のキャラクターのイラストをソフトウェアパッケージの箱に描き、そのキャラクターを誰でも合法的に非商用利用できるようにライセンスを整備した⁴⁾。こうして権利を的確に開放したことで、多様な連鎖反応が起きた。例えば、キャラクターや楽曲のイメージのイラストを描いたり、他のクリエイターが作った楽曲に対して新たな映像を付けて動画にしたり、キャラクターのイラストをスライドショーのように自分の曲に付けて動画にしたりする「二次創作」（派生した著作物の創作）が盛んに取り組まれた。こうした権利を開放する流れは、初音ミク以降に次々と発売された他の歌声合成ソフトウェアでも受け継がれている。

このように、歌声合成の手軽さと特徴的な歌声、キャラクター等が人気を呼び、多くの人々が初音ミクやその周辺の技術・文化・現象に興味をもって多彩な活動をしたことで、未来を切り拓く画期的な素晴らしい現象が起きている⁷⁾。これは「初音ミク現象」とも呼ばれ、今日では、歌声合成技術が普及することは歴史的必然であると確信できる状況となった⁸⁾。

一般の人々が活発に創作活動を繰り広げて内容が生成されていくこのようなメディアは、CGM (Consumer Generated Media, 消費者生成メディア) と呼ばれる。歌声合成に関連した創作を促進し、連鎖反応が盛り上がる場として貢献したのがCGM作品を中心とした動画共有サービス「ニコニコ動画」(http://www.nicovideo.jp/video_top/) である。ここでは、動画の視聴を共通体験としたうえで、作品内の時刻に同期したコメント（時刻同期コメント）等によるメタレベルのコミュニケーションを、視聴者同士で、あるいは動画のクリエイターと視聴者間で楽しむことができる。動画作品を投稿すると、時刻同期コメントが視聴者からフィードバックとして

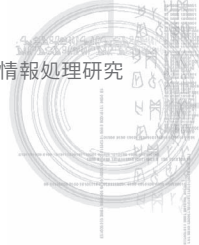
得られ、クリエイターの励みになるため⁹⁾、初音ミク等の楽曲や動画の作品を発表する場として最適であった。

このニコニコ動画の場で、クリエイターと視聴者の創意工夫の相乗効果により多様な二次創作が生まれた。例えば、歌声合成による元楽曲を人間が歌ったり、演奏したり、踊ったりする実写動画や、その実写の踊りをさらにCGキャラクターで再現した動画等も発表されている。こうした創作が創作を呼び起こす連鎖反応（集団的創造現象）は「N次創作」と呼ばれる¹⁰⁾。その本質にはコミュニケーションの要素もあり、「この作品に対して、私はこういう作品を二次創作して応えよう」といった掛け合いのような創作の連鎖が起きている。その過程で複数の作品の良いところが掛け合わされて良質な作品が生まれ、さらに人気が高まっていった。

2. 初音ミクとN次創作に関連した音楽情報処理研究

以上のCGM現象では、日本の技術・社会・文化のすべての強みが結集したと思えるような明るい未来が築かれつつあり、音楽情報処理研究者の立場からの研究開発も活発に取り組まれている。特に、音楽情報処理分野の部分領域である「歌声情報処理」¹¹⁾は以前から研究開発が進んでいたが、初音ミクの登場以降、社会的な関心が高まり、さまざまな大学や企業、研究機関、個人が参入してきている。このほかにも、ニコニコ動画上での創作活動も研究されており、例えば、初音ミク関連のN次創作における協調的創造活動（引用関係や役割等）が分析された¹²⁾。

以下、そうした初音ミクとN次創作に触発されたわれわれの研究事例として、歌声合成技術VocaListener^{13), 14)}と音楽視聴支援サービスSongrium^{15), 16)}を紹介する。



3. お手本歌唱の歌い方をまねる歌声合成技術VocaListener (ばかりす)

VocaListener (ボーカリスナー, 略称「ばかりす」)^{13), 14)} は, 「初音ミク」などの歌声合成ソフトウェアを用いて, 録音された歌声の事例からその歌い方(声の高さと声の大きさ)をまねて歌声が合成できる技術である。人間がお手本を「歌って示す」ことで, より自然なニュアンスで歌声が合成できる。従来は, 人間らしい自然な歌声を合成しようとすると, その細かいニュアンスを表現するために, 楽譜と歌詞を入力した後に, 歌声合成パラメーターをユーザーが人手で長時間調整しなければならないことが多かった。その楽譜入力と調整をVocaListenerにより自動化することで, 調整時間が大幅に減少し, 調整に関する知識をもたないユーザーでも高品質な歌声合成結果を得ることが容易となった(図1)。これによりユーザーは, 合成された歌声によってどのような表現をしたいのか, どのようなメッセージを伝えたいのかに, より注力して歌声が合成できるようになる。

われわれはVocaListenerを実現するために, 反復推定に基づく歌声合成パラメーター推定技術を開発した。この技術によって合成結果の品質が向上し, 歌声合成ソフトウェアやその音源(歌手の声)を切り替えても再調整せずに自動的に歌声が合成できるようになった。その際, 歌詞と歌声の自動対応付け技術によって, 楽譜を一切入力しなくても歌うだけで

合成が可能になった。さらに, ユーザーによる合成結果の微調整を容易にする技術も開発した。

VocaListenerは, 以下の2つの特長をもつ。

- (1) 反復推定に基づく歌声合成パラメーター推定技術によって, 合成結果の品質が向上し, 歌声合成ソフトウェアやその音源(歌手の声)を容易に切り替えて合成可能

VocaListenerを実現するうえで難しいのは, 歌唱事例の歌い方の分析結果(声の高さと大きさ)を, そのまま歌声合成ソフトウェアにパラメーターとして与えても, 音源(歌手の声)によって声の高さと大きさが異なって合成されることである。これは歌声合成ソフトウェアが通常, 多数の短い歌声波形の断片を切り貼りして合成音を出す仕組みをもち, 前後関係にも依存した非線形な動作をするからである。そこでVocaListenerでは, あたかも何度も発声練習するかのように, 合成音を再度取り込んで分析し, 意図したとおりでない部分のパラメーターを補正して再度合成する処理を何度も反復することで, 歌い方を高精度にまねた歌声合成を実現した。これにより, 歌声合成ソフトウェアやその音源(歌手の声)を切り替えても, その新たな声に合わせて自動的に再度パラメーターが調整されるため, 多様な歌声で容易に合成できるようになった。

- (2) 歌詞と歌声の高精度な自動対応付け技術により, 楽譜を一切入力しなくても歌詞のテキストに合わせて歌うだけで合成可能

歌唱事例において, 歌詞のどこをいつ歌っているかを自動的に対応付けることで, 歌詞の各音節の音高(声の高さ)を推定し, 音符化して歌声合成用の楽譜表現を生成可能にした。その自動対応付け技術の精度を向上させるために, 多量の歌声に音節の記号をラベル付けし, 歌声専用音響モデル(歌声特有の音響的な響き方を学習したモデル)を構築した。これにより, ユーザーが音符とその音節を1つひとつ入力してタイミングを調整しなくても, 歌うだけでタイミングが最適に調整された音符列を得ることが

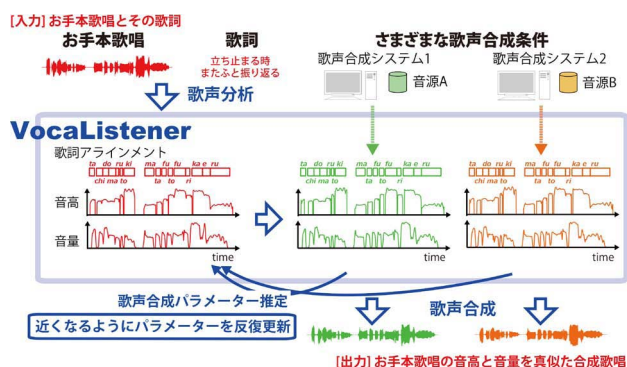


図1 VocaListenerの処理概要

人間の歌声と歌詞を入力して, その歌い方に近くなるように歌声合成パラメーターを反復推定して歌声合成する。

可能になり、利便性が向上した。

VocaListenerは、産業技術総合研究所が単独で研究開発をして2008年に発表した成果であるが、2012年10月19日にヤマハ株式会社から製品名「VOCALOID3 Job Plugin VocaListener」として発売が開始され、誰でも購入すれば利用可能な技術となった。動画1に、VocaListenerのデモンストレーション動画を示す。ほかにも、さまざまな合成結果のデモンストレーション動画が<http://staff.aist.go.jp/t.nakano/VocaListener/index-j.html>で閲覧できる。

4. 音楽コンテンツの関係性を可視化する 音楽視聴支援サービスSongrium (ソングリウム)

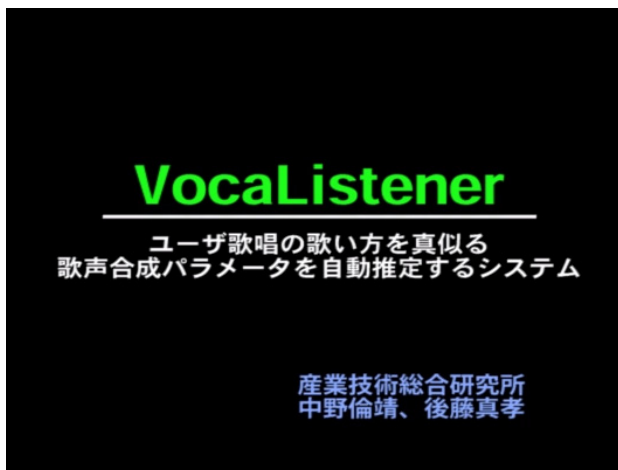
「Songrium (ソングリウム)」(<http://songrium.jp>)^{15), 16)} は、Web上の動画共有サービスで公開されている音楽コンテンツ（オリジナル楽曲を収録した動画や、その「N次創作」の派生作品となる動画）間の関係性を可視化し、さまざまな「つながり」を意識しながら鑑賞できる音楽視聴支援サービスである。2013年8月27日に一般公開し、実証実験を開始しているので、誰でも無料で利用できる。

動画共有サービスの普及により、クリエイターが

創作した膨大な数の音楽コンテンツが公開されているが、ユーザーがそれらの中から関心のある動画を見つける手掛かりとして、音楽コンテンツ同士がもつ関係性、例えばオリジナル楽曲と派生作品との間にある派生関係や、楽曲の類似関係、クリエイターの人間関係など多様な関係性を活用できる音楽視聴支援システムはなかった。

Songriumは、Webマイニング技術および音楽理解技術に基づき、動画共有サービス（ニコニコ動画、YouTube）上の音楽動画60万件以上の関係性を抽出し可視化することができる。これは派生関係を可視化する「惑星ビュー」機能や、楽曲間の関係性をユーザーが自由に追加できる「矢印タグ」機能、曲調や歌声の特徴に基づく楽曲群や派生作品群の可視化機能をもつ音楽視聴支援システムであり、誰でも利用できるWeb上のサービスとして公開した。ユーザーは一連の可視化機能によって明らかになる多様な関係性を手掛かりに楽曲や派生作品と出会うことができる。さらに試聴を容易にするサビ出し機能や、さまざまな形態でSongriumを利用するためのインターフェース（自動連続再生、ブラウザ拡張）も実現した。動画2に、これらの機能を紹介するSongriumのデモンストレーション動画を示す。

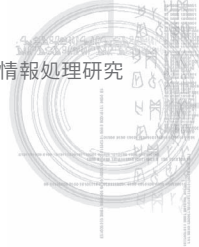
Songriumは、以下の3つの特長をもつ。



動画1 VocaListenerのデモンストレーション動画
(<http://www.youtube.com/watch?v=XpOINZcWkh4>)



動画2 Songriumのデモンストレーション動画
(<http://www.youtube.com/watch?v=9Uc4Tz99INE>)



(1) 音楽コンテンツのつながりを意識した鑑賞を実現する「関係性ブラウジング」

Webマイニング技術により、ニコニコ動画の場合には「初音ミク」などの歌声合成ソフトウェアを用いている約10万件のオリジナル楽曲を発見して登録するとともに、それらと派生関係にある約50万件の作品群を自動的に抽出して登録した。そして、自動抽出された楽曲と派生作品との関係を「惑星ビュー」によって可視化した(図2)。さらにそうした派生関係に加え、オリジナル楽曲間の任意の関係性を「矢印タグ」として可視化することができる。矢印タグはユーザーが誰でも付与して共有することができる。ユーザーがSongriumに登録された楽曲を選べば、関係性が可視化された画面を見ながら、元の動画共有サービス上にある楽曲やその派生作品の動画をストリーミング再生して楽しむことができる。

「惑星ビュー」では、派生作品の種類や視聴回数が色やサイズなどで直感的に表されており、楽曲によって見た目が大きく変わる。このような派生作品群の可視化により、オリジナル楽曲を視聴しているだけでは気付きにくい派生作品が発見できる。さらにオリジナル楽曲がどのような種類の派生作品を生み出

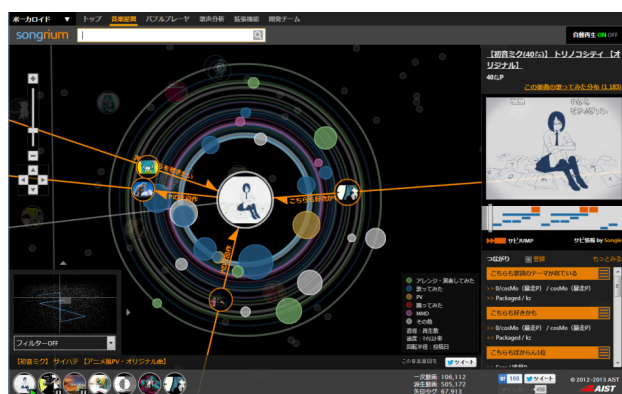


図2 オリジナル楽曲の「惑星ビュー」と「矢印タグ」の表示例
中心の丸いサムネイル画像がオリジナル楽曲を、その周りを回転するカラフルな複数の円形図形が派生作品群を示している。楽曲を中心とした同心円はそれぞれが派生作品の回転軌道を表している。回転半径は派生作品の投稿時期を示しており、新しい派生作品ほど外側にある。円形図形の大きさは派生作品の再生回数、色は派生作品の種類(例えば青色は楽曲を歌唱した派生作品、赤色は楽曲に合わせて踊った派生作品)を意味する。中央の楽曲に向けて伸びている直線は矢印タグを示しており、矢印タグ上にある丸いサムネイル画像はつながったもう一方の楽曲を示している。この矢印タグをたどることで次々に楽曲を視聴できる。

してきたかも知ることができる。

「矢印タグ」は、ある楽曲が別の楽曲に対してどのような関係にあるのかをテキストで自由に記述して共有する仕組みである。一般的なソーシャルタギングではコンテンツ単体に対してタグ付けするのに対し、矢印タグは2つのコンテンツ間の関係に対してタグ付けできる特長をもつ。同一作者の次回作やデビュー作のようなWebマイニング技術で発見できる関係については、矢印タグとして自動生成される。それに加えて、ユーザーが自由に矢印タグを追加できるインターフェースによって、自動処理では発見が難しい楽曲間の多様な関係も共有することができる。

(2) 楽曲の中身を自動解析した結果を利用した音楽星図、サビ出し機能などによる鑑賞支援

音楽理解技術によって楽曲の中身を自動的に解析した結果を利用した以下の鑑賞支援機能により、さらに豊かな鑑賞を可能にする。

(a) 音楽星図：楽曲の曲調を分析し、2次元平面上で曲調の似た曲を近くに配置することで、Songriumにすでに登録された音楽動画60万件のうちオリジナル楽曲にあたる10万曲を俯瞰できる「音楽星図」を実現した(図3)。曲調が似た曲を発見したいときには、音楽星図上で近くに存在する楽曲をクリックすると、それが中央に

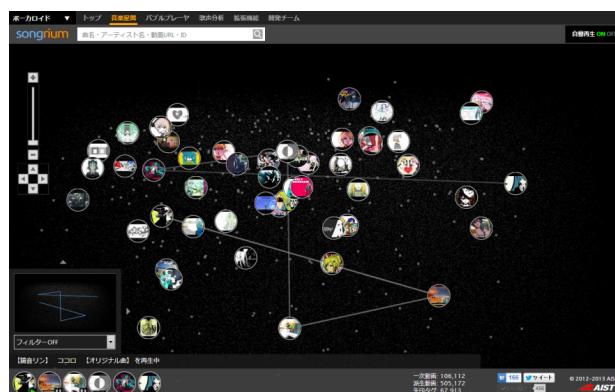


図3 楽曲の曲調を自動解析して「音楽星図」として可視化した例
約10万曲のオリジナル楽曲が、曲調の近さに基づいて2次元平面上に配置されている。地図サービスのように、マウス操作でスクロールやズームングをして楽曲を探索できる。再生回数の多い楽曲は丸いサムネイル画像で表示されている。

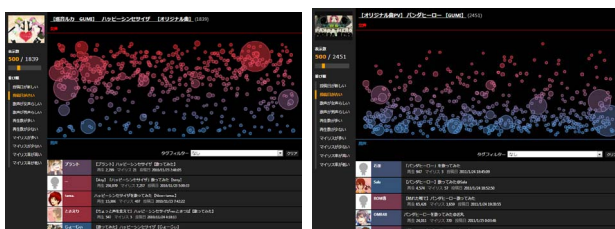
移動してきて視聴できる。

(b) 歌声の男女度：同一楽曲をさまざまな歌手が歌唱した派生作品群（以下、歌唱動画）に対して、歌声の音響的特徴から男女度（男声らしいか女声らしいかを示す指標）を自動推定し、その結果を可視化する機能を実現した（図4）。この機能によって、男女度の違いを俯瞰しながらさまざまな歌唱動画を聴き比べることができる。ただし、ニコニコ動画上の音楽コンテンツのみに対応している。

(c) サビ出し機能：楽曲中で一番代表的な盛り上がる主題の部分である「サビ」を容易に見つけて試聴できる「サビ出し機能」を実現した（図5）。これはわれわれが研究開発中の能動的音楽鑑賞サービス「Songle（ソングル）」（<http://songle.jp>）¹⁷⁾の「サビ出し機能」と連携している。自動解析したサビの位置の誤りにユーザーが気付いた場合は、それをSongle上のインターフェースで訂正することで、Songrium上の表示も正しくなる。

(3) 音楽視聴支援サービスを多様なインターフェースで活用

Webマイニング技術、音楽理解技術、さらにユーザー入力によって得られた関係性をより多くのユーザーに利用してもらうために、多様なインターフェー



女声らしさの高い派生作品が多いオリジナル楽曲

男声らしさの高い派生作品が多いオリジナル楽曲

図4 楽曲を歌唱した派生作品における歌声の男女度分布の表示例
歌唱動画の派生作品は、動画共有サービス上で作品数も視聴回数も多く、人気が高い。歌唱動画同士では曲自体には違いがないので、声質や歌い方の違いが重要となる。画面中央の赤や青の円形図形は歌唱動画を表しており、上にあるほど歌声の女声らしさが高く、下にあるほど男声らしさが高く、と自動推定されている。このように歌声の声質を可視化することで、ユーザーは歌声の違いを意識しながらさまざまな歌唱動画を聴き比べやすくなる。

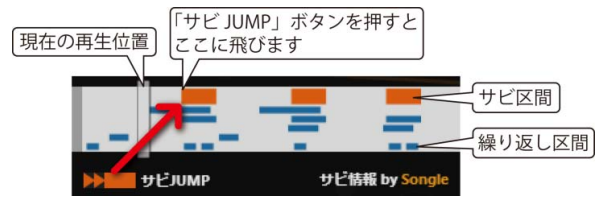


図5 サビ出し機能の使用例

楽曲構造を可視化することで、任意のサビ区間や繰り返し区間を押してジャンプしながら効率的に試聴できる。左下の「サビ JUMP」ボタンを押して「次のサビ区間の頭出し再生」もできる。

スを用意した。

(a) 自動連続再生インターフェース「バブルプレーヤ」：ユーザーが指定した期間に投稿された楽曲群を自動連続再生できる。投稿日時順に楽曲が次々と出現して増えていく様子をアニメーション表示しつつ、再生回数などユーザーが指定した条件を満たした楽曲群の一部分を順次再生して紹介することで、1つのムービーのように視聴することができる（図6）。これにより動画共有サービス上のトレンドの変遷が容易に俯瞰できる。ただし、ニコニコ動画上の音楽コンテンツのみに対応している。

(b) ブラウザ拡張「Songrium Extension」：「ブラウザ拡張」機能に対応したWebブラウザを使用しているユーザーに対して、Songrium専用のブラウザ拡張を提供する。これによりSongriumにアクセスしなくても、ニコニコ動画やYouTube上でオリジナル楽曲や派生作品を閲覧するだけで、直接、サビ出し機能等のSongriumの機能の一部を使用できる。

5. 自動創作と自動鑑賞の可能性

冒頭で述べたように、「人間の歌声でなければ聴く価値がない」という旧来の価値観が打破され、「合成された歌声がメインボーカルの楽曲を積極的に楽しむ文化」が誕生したのは衝撃的であった。それと同時に、音楽情報処理の発展が今後も別の既存の価値観を打破し、さらに新たな文化を生む可能性を感じ

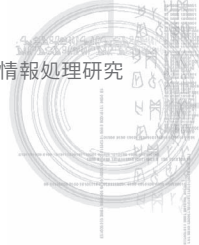


図6 指定期間のオリジナル楽曲群をムービーのように鑑賞する「パブルプレーヤ」の画面例

ユーザーが期間を指定すると、動画共有サービスにその間に投稿された楽曲群が投稿日時順に現れ、投稿された楽曲群の変遷を俯瞰できる。楽曲はそれぞれ1つのパブル（着色された円形図形）で表されており、楽曲の投稿に合わせて次々とパブルが増えて集まっていくアニメーションが表示される。パブルの色は使用されている歌声合成ソフトウェアの歌声ライブラリの種類を、大きさは再生回数を表している。

させてくれた。では、より進んだ1つの究極的な可能性として、「計算機が自動的に創作して、計算機が自動的に鑑賞する未来」はありえるのだろうか。音楽がもたらす感動の本質を考察するためにその議論を試みる。

5.1 自動創作

まず、「自動創作」から考える。例えば、さまざまな未来の可能性の1つとして、楽曲や歌詞すら技術の力によって全自動で生成され、それを人々が積極的に楽しむ文化も誕生するかもしれない。すでにさまざまな研究者が自動作曲に向けた研究開発に取り組んでいるものの、現時点では、われわれ自身も「人間の作品だからこそ楽しめる」という考えを捨てにくい。しかし、「人間の歌声でなければ聴く価値がない」という旧来の価値観が打破された以上、「人間の作品でなければ聴く価値がない」という価値観が打破されない保証はない。

すでに、動画共有サービスで歌声合成による作品

を楽しむ文化においては、作者が人間であるか、計算機であるかを、音以外の証拠からつかむことは不可能になった。これまでも、歌声以外の楽器音は電子楽器で合成可能になり、非専門家には区別のつかない高いクオリティとなって音楽制作で用いられてきたが、少なくともメインボーカルの歌声からそこに人間の介在が確信できた。しかし、メインボーカルとして歌声合成が受け入れられた現在、もはや、完全に全自動で歌が作られて動画共有サービスに投稿されても、気付くことができない環境でわれわれは音楽を楽しんでいるのである。この環境では、計算機が音楽（歌詞、歌唱、伴奏）を自動生成して投稿した作品であっても、人間が気付かずに再生してコメントを付ければ、計算機側は再生数やコメントの内容を自動理解して改善の手掛かりとして活用できる。改善しながら何度も作品投稿を繰り返すことで、よりよい作品が自動生成できるようになる潜在的な可能性を秘めている。

さらに、現在の技術ではその実現可能性はまだ低くても、計算機が本当の意味で歌い始めたら、つまり、何らかの感情と解釈されうる状態をもち、それを自己主張として歌で表現し始めたとしたら、われわれ人間はどのように思うのだろうか。あるいは、もし「音楽の自動生成」という言葉に抵抗はあっても、その自動生成の仕組みが実は、人間が作品として発表している膨大なCGM作品のN次創作であり、人間と計算機との「仮想合作」と呼べる状態だったらどうか。つまり、膨大なユーザーがCGM文化の中で創り出す楽曲が膨大な曲数（例えば数百万曲）に達し、それらの楽曲に対する自動音楽理解結果に基づいて新たな素晴らしい作品が自動生成されたときに、われわれ人間はどのように思うのだろうか。

5.2 自動鑑賞

以上の議論では、少なくとも聴衆としては人間を前提にしていた。しかし、聴衆が計算機という可能性もあるのだろうか。「自動鑑賞」が可能になった未

来を考える。

現在の計算機による音楽の自動理解の研究では、音楽音響信号中に表現されるさまざまな音楽的要素や現象を記述したり、ジャンル分類やタギング、感情等の事前に学習したラベルを付与したりする取り組みが多い。動画共有サービス「ニコニコ動画」の時刻同期コメントと音楽との関係を機械学習して、音楽に対してコメントを自動生成するシステム「MusicCommentator」¹⁸⁾も実現された。しかし、音楽を人間と同じように好みや価値判断までもなう形で「鑑賞」するような技術は、実現されていない。このように現在の技術ではその実現可能性はまだ低い、もしそれが可能になったとしたら、われわれ人間はどのように思うのだろうか。

人間は、自分が音楽を創作した後に、その作品を「人間が聴いてくれる」ことに価値を感じる。作品を表現すること自体に価値を感じて誰にも聴いてもらわない場合もあり得るが、通常は、作品に表現した何らかのメッセージや気持ちを他の人間が受けとめてくれたり、それがコミュニケーション的な役割を果たしてくれたりすることを期待する。したがって、これも自動創作同様、現時点では、人間ではなく「計算機が聴いてくれる」としても、そこに価値は感じにくい。しかし、これも「人間の聴衆でなければ聴いてもらう価値がない」という旧来の価値観として打破される可能性がないとは言い切れない。

さらには、「自動創作」と「自動鑑賞」がともに技術的に可能になった未来においては、計算機が創作して、計算機が鑑賞する状況もあり得る。人間が鑑賞するのが前提でなければ、聴覚器官や人間の理解力の制約を受けなくなり、テンポが極めて速くなったり音響表現が極めて複雑になったりするなど、音楽表現が変容していく可能性すらある。それはSF的で興味深い、現実には、そうした未来でも人間が創作し、人間が鑑賞する世界は続くはずである。そうであれば、これまで同様、新たな技術の登場が新たな音楽表現を生んでいく変容は起きつつも、人間

も受容して楽しめる範囲での変容の可能性の方が高いであろう。

5.3 音楽がもたらす感動とは

自動創作のような技術によって創作された音楽は、どういう感動をもたらすのであろうか。それが純粋に素晴らしいければ受け入れられるのか、それとも潜在的にはいくら感動できる音楽であったとしても、人間が創作していないと聴衆が知ってしまった段階で、感動できなくなるのだろうか。

コンテンツ全般（ここでは音楽）がもたらす感動には、少なくとも以下の3つの要因が絡み合っているという仮説をわれわれは考えている¹⁹⁾。

(i) コンテンツ自体の感動

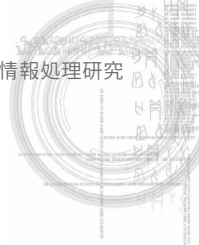
コンテンツそのものによって純粋に引き起こされる感動である。そのコンテンツが創作された過程や、それが社会的、文化的にどういう意味をもっているのかには無関係に感じられる感動といえる。

(ii) オリンピック的感動

同じ生身の人間が創作しているとわかっているからこそその感動である。例えば、いくら早弾きのすごい演奏を含む音楽であっても、それが計算機によって自動演奏されていると知っていれば感動しないが、人間が目の前で演奏していれば感動する。それは、オリンピックで選手が走っている横を、自動車であってゴールに到着しても感動しないのと同様であり、オリンピックでは自分と同じ人間がその行為をしている点、感動を生んでいる。

(iii) 文脈的感動

コンテンツがおかれた文脈を知っていることにより生まれる感動である。同じコンテンツであっても、それが誕生した社会的、文化的、個人的背景を知っていることで、人々は感動することがある。この状況でこういう表現をしたコンテンツだから、あのコンテンツの影響を受けて生まれたコンテンツだから、あの人がこうしたい



という理由で創作したコンテンツだから、感動する場合がある。上記の「オリンピック的感動」も、「創作過程という文脈」に特化した文脈的感動という解釈もできる。

本稿では上記の3つを取り上げたが、感動には他の要因もある。例えば、ほかの人との共同体験であることによって得られる感動（この人と一緒にこの曲が聴けて嬉しいというような感動）などがあげられる。文脈的感動に近いが、文脈がコンテンツ側でなく、鑑賞者側にある。

従来は、(i)～(iii)の感動は切り分けることは困難であった。コンテンツが発表されるときには、誰がどういう形で発表しているかが見えやすかったからである。しかし、現在では、例えば動画共有サービスに創作した音楽の動画を投稿した場合、(i)の純粋なコンテンツ自体の感動があれば、高い評価を受けられる可能性がある（ただし、埋もれて高く評価されない場合も多い）。さらに、コンテンツ創作過程も公開すれば、(ii)のオリンピック的感動の可能性が増えるし、ある文脈に意図的に位置付けて発表すれば、(iii)の文脈的感動を引き起こせる可能性がある。

では今後、技術がより一層進歩して、自動創作が可能になったり、音楽の創作や演奏などが手軽になって誰でも可能になったりしたときに、人はどう感動するのだろうか。もし純粋に(i)のコンテンツ自体の感動を感じられるならば、その観点では、何ら変わらない感動が得られるであろう。逆に、もしそこに(ii)のオリンピック的感動が不可欠なのだとすると、感動できなくなる可能性がある。それが非常に簡単に創作されたと知ってしまうと、高く評価しにくい気持ちが生じる場合があるからである。その状況下でも、例えば、初音ミクのようなキャラクター（擬人化した仮想の身体性をもつ存在）がコンテンツを創作したと受け止められれば、(iii)の文脈的感動と(i)のコンテンツ自体の感動が得られるのかもしれない。

3章で述べた、人間が歌うだけで手軽に自然な歌声が合成できるわれわれの歌声合成技術「VocaListener

(ばかりす)」でも類似した議論はあった。本技術による自然な歌声合成結果の楽曲に対して、(i)「コンテンツ自体の感動」で純粋に高く評価する人々がいた一方で、それが手作業によるパラメーター調整でなく技術の力だと知ったことによって、(ii)「オリンピック的感動」が失われて低い評価に転じる人々もいたのは興味深い（ただし、実際にはこの技術を使った場合でも、クリエイターの裁量で表現を調整したり工夫したりできる余地が多く、いずれ理解が進めば(ii)の感動も広まる可能性はある）。また、われわれの歌声合成文化の発展を真に願う気持ちや長年の取り組みを知っている方々からは、(iii)「文脈的感動」を感じて高い評価をいただく場合もあり得る。

6. おわりに

初音ミク現象やN次創作を引き起こすCGM現象は、技術の発展が新たな価値観や文化を生み出した事例として大きなインパクトを与えてきたが、現在も多様性を増しながらさまざまな形で広がり続けている。技術や文化が発展していくうえで、多様であること、多様な価値観が共存共栄することは重要である。

例えば、歌声合成技術が発展すると「人間の歌手はいらなくなるのか」についてしばしば議論されるが、われわれ人類が歌うことをやめない以上、歌手がいらなくなるとは思えない。人は、強制されて歌うわけではなく歌いたいから歌っているし、聴き手も、この人（好きな歌手、友だち、自分の子供）が歌うから、と思って聴くからである。単に歌声合成という選択肢が増えて表現が多様になるだけであり、すべてが歌声合成になってしまって多様性が失われるのは望ましくない。

合成された歌声がもつ表現力においても多様性は重要であり、3章で紹介したVocaListener（ばかりす）を使った自然な合成結果よりも、使わない人工的な合成結果の方が、むしろ歌声合成らしさが強くて人間の歌手とは違う表現として好ましいという価値観

をもつ人々もいる。こうした自然さを増す技術もあくまで選択肢を増やす手段なのであって、合成された歌声すべてが自然になってしまうのは、むしろ不自然なのである。初音ミクが発端となったCGM現象は、海外にもより一層広がりがつつあり、今後、世界中の多様な価値観・文化・表現・技術を尊重し合うことで、さらなる未来が切り拓かれていく。

初音ミク現象には、人と人をつなげるハブ (hub) の効果があり、異分野にいる人々が、初音ミクや歌声合成をキーワードに次々とつながってきた。クリエイター同士も、N次創作のように多くのコンテンツが絡み合う過程で交流が促進され、視聴者を交えたコミュニケーションも活発である。4章で紹介したSongriumが、そうした「つながり」をより深めていく場としても機能することをわれわれは目指している。学術的な研究開発だけでなく、周辺のソフトウェア・ハードウェア開発や動画ランキング制作なども活発で、さまざまなコラボレーションが生まれる土壌となっている。

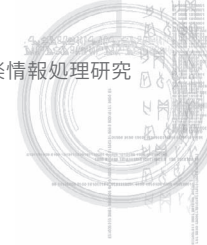
これは技術の力で切り拓かれつつある新たな世界であり、単に「歌やキャラクターを好む人々のための趣味的な世界」だとみなすと、この本質を見誤る。実際、本現象に注目している研究者には、これまでのキャラクターを中心とした文化にはなじみが薄く、初音ミクで初めて議論している人々も多い。日本が海外に対して圧倒的に先行しているこの日本発の現象を今後どう発展させていけるかは、1人ひとりの行動にかかっている。本稿をきっかけに、より多くの人々がこの技術の力で切り拓かれつつある新たな世界に興味をもち、さまざまな形でそのさらなる発展に関わっていただければと願っている。

謝辞

SongriumのインターフェースデザインとWebサービスの実装を担当していただいた石田啓介氏に感謝する。本研究の一部は科学技術振興機構 (JST) CRESTの支援を受けた。

参考文献

- 1) Kenmochi, Hideki; Ohshita, Hayato. Vocaloid — Commercial Singing Synthesizer Based on Sample Concatenation, in Proc. of Interspeech 2007. 2007, p. 4009-4010.
- 2) 剣持秀紀. 歌声合成の過去・現在・未来:「使える」歌声合成のためには. 情報処理 (情報処理学会誌). 2012, vol. 53, no. 5, p. 472-476.
- 3) 剣持秀紀. 歌声合成システムVOCALOIDの開発. システム／制御／情報 (システム制御情報学会誌). 2012, vol. 56, no. 5, p. 244-248.
- 4) 伊藤博之. 初音ミク as an interface. 情報処理 (情報処理学会誌). 2012, vol. 53, no. 5, p. 477-482.
- 5) 剣持秀紀. 歌声合成技術の動向 — 「初音ミク」を支える技術 —. 日本音響学会誌. 2011, vol. 67, no. 1, p. 46-50.
- 6) 後藤真孝. 「初音ミク」はなぜ注目されているのか. 電気学会誌. 2012, vol. 132, no. 9, p. 630-633.
- 7) 後藤真孝. 初音ミク, ニコニコ動画, ピアプロが切り拓いたCGM現象. 情報処理 (情報処理学会誌). 2012, vol. 53, no. 5, p. 466-471.
- 8) 後藤真孝. 音楽情報学. 情報処理 (情報処理学会誌). 2010, vol. 51, no. 6, p. 661-668.
- 9) 戀塚昭彦. ニコニコ動画の創造性: 動画コミュニティサービス「ニコニコ動画」の5年間. 情報処理 (情報処理学会誌). 2012, vol. 53, no. 5, p. 483-488.



- 10) 濱野智史. ニコニコ動画はいかなる点で特異なのか:「疑似同期」「N次創作」「Fluxonomy (フラクソノミー)」。情報処理 (情報処理学会誌). 2012, vol. 53, no. 5, p. 489-494.
- 11) 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. 2008, vol. 64, no. 10, p. 616-623.
- 12) 濱崎雅弘, 武田英明, 西村拓一. 動画共有サイトにおける大規模な協調的創造活動の創発のネットワーク分析 — ニコニコ動画における初音ミク動画コミュニティを対象として. 人工知能学会誌. 2010, vol. 25, no. 1, p. 157-167.
- 13) 中野倫靖, 後藤真孝. VocaListener: ユーザ歌唱の音高および音量を真似る歌声合成システム. 情報処理学会論文誌. 2011, vol. 52, no. 12, p. 3853-3867.
- 14) 後藤真孝, 中野倫靖, 梶田秀司, 松坂要佐, 中岡慎一郎, 横井一仁. 人間の歌い方を真似る歌声合成システム VocaListenerとロボット顔動作生成システムVocaWatcher. システム／制御／情報 (システム制御情報学会誌). 2012, vol. 56, no. 5, p. 249-255.
- 15) 濱崎雅弘, 後藤真孝. Songrium: 関係性に基づいて音楽星図を渡り歩く音楽視聴支援サービス. WISS 2012 論文集. 2012.
- 16) Hamasaki, Masahiro; Goto, Masataka. Songrium: A Music Browsing Assistance Service Based on Visualization of Massive Open Collaboration Within Music Content Creation Community, in Proc. of the 9th International Symposium on Open Collaboration (WikiSym + OpenSym 2013). 2013, p. 1-10.
- 17) 後藤真孝, 吉井和佳, 藤原弘将, Matthias Mauch, 中野倫靖. Songle: 音楽音響信号理解技術とユーザによる誤り訂正に基づく能動的音楽鑑賞サービス. 情報処理学会論文誌. 2013, vol. 54, no. 4, p. 1363-1372.
- 18) Yoshii, Kazuyoshi; Goto, Masataka. MusicCommentator: Generating Comments Synchronized with Musical Audio Signals by a Joint Probabilistic Model of Acoustic and Textual Features, in Proc. of ICEC 2009. 2009, p. 85-97.
- 19) 後藤真孝. 招待講演「未来を切り拓く音楽情報処理」. 情報処理学会研究報告 (音楽情報科学). 2013, vol. 2013-MUS-99, no. 33, p. 1-9.

Author Abstract

In this paper we introduce two examples of music information processing research related to "Hatsune Miku" and N-th Order Derivative Creation: a singing synthesis technology "VocaListener" and a music browsing assistance service "Songrium". VocaListener is a technology that uses singing synthesis software like "Hatsune Miku" to synthesize natural singing voices by imitating the pitch and dynamics of recorded examples of human singing voices. Songrium is a music browsing assistance service that visualizes various kinds of relations, such as relations among both original songs and their derivative works generated from the N-th Order Derivative Creation. This paper then examines future possibilities of automatic creation and automatic appreciation and also discusses essential factors of emotional, touching experiences elicited by music.

Key words

singing synthesis, Hatsune Miku, CGM, N-th Order Derivative Creation, music information processing, singing information processing