

未来を切り拓く音楽情報処理

後藤 真孝^{1,a)}

概要：音楽の楽しみ方の未来は、音楽情報処理技術によってどのように切り拓かれていくのだろうか。世の中のすべての音楽はデジタル化される宿命にあり、いつでもどこでも好きな音楽を大量に聴ける時代が到来した。従来は、音楽は受身で聴くことが多かったのに対し、これからは、もっと自分の好みに合わせて自在に音楽を楽しむことが可能になっていく。本招待講演では、音楽を鑑賞する場面、制作する場面の両方において、音楽情報処理技術がどのように新たな可能性を切り拓いていくのかを議論する。

1. はじめに

2013年度に情報処理学会 音楽情報科学研究会 (SIGMUS) は発足 20 周年を迎えたが、過去 20 年間で音楽情報処理分野は大きく飛躍し、学術的、産業的、社会的、文化的に重要性を増し、より一層注目されるようになってきた。音楽情報科学研究会は、1985 年に任意団体として出発し、1989 年からの情報処理学会研究グループを経て、1993 年から情報処理学会の正式な研究会となった [1]。現在の活況は、その当時から考えたら夢のようである。1957 年の「イリアック組曲」のように計算機（本稿では「コンピュータ」の意味で用いる）の発明当初から自動作曲への応用が試みられる等、音楽情報処理分野の歴史は長く、その研究成果は、音楽制作で不可欠となったシンセサイザや音楽配信等、社会に広く普及してきた。

特に 2000 年代に入ってから、すべての音楽がデジタル化され、配信・検索・共有・創作・発信されることを人々が実感として理解した結果、重要な研究分野であることが広く認知され、世界中で新たな研究者が参入し続けている。単調増加する膨大な楽曲をどう扱うかという問題意識から、音楽情報検索・推薦、新たな音楽インタフェース等に関する研究が活発になされてきた。

学術的にも「複数の音が相互に関係し合いながら時間的な構造を形成して内容を伝える信号」の理解という未解決で本質的な課題を持つことが、多くの研究者に注目されている 1 つの理由である。他にも、未解決で魅力的な課題はまだまだ手つかずの状況であり、研究テーマの宝庫である。この音楽情報処理分野が学術的にいかに多方面から注

目されてきたかは、様々な学会の学会誌で企画された特集号 [2], [3], [4], [5], [6] や解説 [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21] の多さだけを見てもわかる。関連成果が発表される国際会議・ワークショップも増え続け、分野の重要性と将来性から、大型予算による研究プロジェクトが欧米や日本を中心に次々と生まれている。

デジタル音楽コンテンツの普及と共に、音楽情報処理技術は、既に様々な形で産業応用が拡大して社会に貢献してきた。2007 年以降は、歌声合成をはじめとする「歌声情報処理」[9] が特に注目を集めている。最も衝撃的なのは、「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とである。これは歌声合成技術「VOCALOID2」[20] に基づく歌声合成ソフトウェア「初音ミク」[18], [21] が発売される 2007 年以前には考えられなかった快挙である。それと同時に、今後も音楽情報処理の発展が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させてくれた。

本稿では、こうした音楽情報処理が重要性を増し続けてきた状況を踏まえ、音楽情報処理がより一層学術的に進展し、社会に貢献し、未来を切り拓いていくことを願って、様々な仮説やグランドチャレンジに関する筆者の考えを述べていく。既に解説「音楽情報学」[19] でも、5 つのグランドチャレンジとして、「個人ごとに最適な音楽を提供できるか」、「音楽の流行を予測できるか」、「人間と音楽との関係をより豊かにできるか」、「音楽自体に新たな進化をもたらすか」、「環境問題、エネルギー問題に寄与できるか」について野心的な議論を展開した。本稿では、より一層議論の余地が大きい話題を意図的に取り上げ、現状の技術か

¹ 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)

^{a)} m.goto [at] aist.go.jp

ら予想される次の段階というよりは、既存の価値観にとらわれないためにも、現時点では仮説のみに基づいていたり技術的な実現可能性が低かったりする未来予想を議論することで、コミュニティ全体が新たな発想を広げる一助となることを狙っている。したがって下記は、研究分野全体としてコンセンサスを得た内容ではない点に注意が必要であるが、音楽情報科学研究会 20 周年という機会に刺激的な議論をしてみたい。

2. 音楽情報処理が切り拓く未来

音楽情報処理が可能にする未来には様々な可能性があり、それぞれの研究者が自分が信じる未来を目指して研究開発を進めることで、未来が切り拓かれていく。そうした音楽情報処理における多様な研究アプローチを分類する観点として、解説「音楽情報学」[19]では、

- 「認識・理解」と「生成」
- 「信号処理」と「記号処理」
- 「リアルタイム・インタラクティブ」と「非リアルタイム・非インタラクティブ」
- 「システムを作る」と「人間を知る」

を挙げた。他にも、

- 「一曲を扱う」と「楽曲の集合を扱う」
- 「楽曲そのものを扱う」と「メタ情報を扱う」
- 「音楽演奏によるコミュニケーション」と「音楽を題材としたメタレベルでのコミュニケーション」
- 「プロ向け」と「アマチュア向け（しろうと向け）」
- 「全自動」と「半自動（手動）」

など、様々な観点が考えられる。今後の研究分野の方向性を考えるときに、上記の最後の観点に関連した以下の二つの研究アプローチを意識することが重要になると筆者は考えている。

(1) 人間の能力を拡張する研究（「音楽能力支援」の実現）

音楽家の能力を身につけられるようにトレーニングを支援する研究だけでなく、人間の能力の一時的な拡張を支援する研究も含まれる。例えば、作曲ができない人が、作曲支援システムを利用して楽曲制作したり、歌が上手でない人が、歌声合成システムを利用してメインボーカルを制作したりすることが考えられる。自分はある楽器が演奏できなくても、その楽器の音を容易に出すことができる電子楽器（シンセサイザ）も、技術の力による人間の能力を拡張といえる。「音楽能力支援」については、3 章でさらに議論を深めていく。

(2) 音楽家の能力の一部を全自動で可能にする研究（「仮想の音楽家」の実現）

人間の音楽家が持つ様々な能力を計算機上で実現しようとする研究である。自動作曲、自動編曲、自動作詞、自動演奏（標準 MIDI ファイル (SMF) 再生、即興演奏、演奏の表情付け）、自動伴奏（合奏）、自動読譜（楽

譜認識）、自動採譜のように「自動〇〇」と呼ばれる代表的な研究テーマがある。このアプローチでは、人間の能力を人間らしく真似できるように実現する場合と、人間の能力に拘らずにそれを越えた機能を実現する場合とがある。例えば、自動演奏の場合には、「楽器を自動制御する」あるいは「SMF を自動再生する」という意味においては、「人間が弾けるかどうか」という人間の能力に拘らずに、むしろ人間には弾けない表現が多用されている。一方、自動演奏には、ジャムセッションのような即興演奏を自動生成したり、演奏表情付けコンテスト「Rencon」[22]のように楽譜をいかに人間らしく演奏するかを追求したりする取り組みもある。この仮想の音楽家の方向性のさらに先を考える議論は、4 章で引き続き展開する。

実際には、(1) と (2) はどこにフォーカスを当てているかの違いに過ぎず、連続的に様々な形態があり得る。(2)「仮想の音楽家」で実現した能力が、それを持たない人間を支援するための目的で使われれば、(1)「音楽能力支援」の技術になる。例えば、歌詞を入れて選択肢を選ぶと自動作曲できるシステム「Orpheus」[23], [24] は、自動作曲としては (2) に位置づけられるが、人間が作曲する能力を拡張する技術とみなせば (1) に位置づけられる。

(1) と (2) を考えるとき、どの程度人間が関与すれば、単なる全自動ではなく自分が支援されたと感じられるのか、つまり拡張された自分の能力で自己表現できたと感じられるのかは興味深い。ボタンを一つ押すだけでは、いくら創作支援だと言われても、生成結果が自分の表現とは思えないからである。

3. 人間の能力の拡張

2 章の「(1) 人間の能力を拡張する研究（「音楽能力支援」の実現）」で述べたように、人間の能力には限界があり、それを技術の力で拡張し、限界を突破する様々なアプローチが取り組まれてきた。そうした能力の拡張にはいくつかの方向性が考えられるが、筆者は下記のように分類している。

- (a) トレーニング：人間の能力自体を向上させるアプローチ
- (b) 即効性のある支援：人間の能力を技術の力で一時的に向上させ支援するアプローチ

(b-1) 誰でも音楽家：できない人を可能にする支援

(b-2) 超絶能力を持つ音楽家：できる人をさらにすぐする支援

(a) の「トレーニング」は、人間自身が自らの能力を持続的あるいは恒久的に向上できるように、技術的に支援する研究である。歌唱のトレーニング [25], [26] や演奏のトレーニング [27], [28]、絶対音感のトレーニングなどが取り組まれてきた。技術的な支援がなくても、地道な訓練によって人間は能力を向上することはできるが、より効率的で、よ

り適切に能力を向上できる支援をそうした研究では目指している。

一方、(b)の「即効性のある支援」では、人間の能力自体は変わらなくても、技術的に支援されている間だけ、何らかの音楽的能力が拡張された状態となるようにする研究である。つまり、人間の能力の向上を待たずに、その技術を使っている間だけ可能なことを増やす。誰がどこまで可能になるのかに応じて、ここではさらに(b-1)「誰でも音楽家」と(b-2)「超絶能力を持つ音楽家」の二つに分類して議論を進める。

(b-1)の「誰でも音楽家」は、主に音楽的な訓練を受けていない一般の人々をユーザとして想定し、技術の力で通常だったら困難なことを可能にする研究である。ユーザの意図を何らかの形で反映できる作曲支援 [23], [24] や、楽器が演奏できなくても計算機上の操作で思い通りの演奏が可能な演奏支援 [29], 歌唱を合成したり加工したりする歌唱支援 [20], [21], [30] 等が取り込まれている。訓練を受けた音楽家でも、自分が持っていない能力の支援を受ける場合は、これに該当する。

一方、(b-2)の「超絶能力を持つ音楽家」は、音楽的に訓練を受けた音楽家をユーザとして想定し、通常人間だったら困難な動作や表現を可能にする研究である。計算機の支援で能力を拡張された人間が、どこまで新たな表現を切り拓けるかが追求される。「超絶演奏の支援」と考えれば、電子楽器の制御による高速で機械的な演奏はこれに該当するが、人間の身体を物理的に拡張する等、潜在的には様々な可能性を秘めている。

実際には、(b)の「即効性のある支援」を受けている間に、(a)の「トレーニング」の効果もあったり、(b-1)と(b-2)の中間的な支援もあったりするので、これも連続的に様々な形態があり得る。

以上の例では、音楽創作や演奏等の「生成」に関する人間の能力の拡張を例にした説明が多かったが、以下では、音楽鑑賞等の「認識・理解」に関する人間の能力の拡張の例として、筆者らが取り組んでいる「音楽理解力拡張インタフェース」(3.1章)と「OngaCREST プロジェクト: 類似度を可視化する情報環境」(3.2章)を紹介する。これら以外にも、様々な研究者が人間の能力の拡張に関連した研究をしている。

3.1 音楽理解力拡張インタフェース

(b)の「即効性のある支援」の、(b-1)の「誰でも音楽家: できない人を可能にする支援」の例として、筆者が「音楽理解力拡張インタフェース」[31], [32], [33]と名付けた、「音楽を聴く力」を拡張する支援を目指した研究アプローチを紹介する。

音楽家の能力として、「音楽を創作する力」や「音楽を演奏する力」はよく議論されるが、音楽家でないエンドユー

ザの能力としての「音楽を聴く力」はまだ充分には議論されていない。ここでの本質的問題は、「他の人が音楽をどう理解しているのか」、「自分が音楽を聴いているときに何がわかっていないのか」を知ることが難しいことにある。他の人が楽曲中の何の要素をどう聴いているかを直接観測できないため、間接的に知るしかないが、必ずしも適切に言語化されない。そして、ユーザが「音楽を聴く力」を向上させたい、音楽をより深く理解したいと思っても、自分が何がわかっていないのかがわからないため、それを可能にする方法は自明ではない。

3.1.1 音楽理解力拡張インタフェースの具体例

そこで我々は、エンドユーザが音楽をより深く理解できるようにする研究アプローチとして、音楽の理解力を拡張してくれる「音楽理解力拡張インタフェース」(Augmented Music-Understanding Interfaces) [31], [32], [33]を提案してきた。下記がその具体例である。

- SmartMusicKIOSK: サビ出し機能付き音楽試聴機 [34], [35]

楽曲構造を利用することで、ポピュラー音楽の楽曲内部のブラウジングや楽曲の試聴を容易にする音楽再生インタフェースである。楽曲構造や音楽家の意図を理解する力を拡張してくれる。

例えば、普段、楽曲の構造を意識せずに音楽を聴いている人が、それを意識して聴ききっかけとなりえる。また、ユーザがサビだけを続けて聴くことで、サビの繰り返しごとに歌詞やアレンジが変化していく様子(そこに込められた音楽家の意図)を、よりの確に把握できる可能性もある。

- LyricSynchronizer: 歌詞と楽曲との自動的な時刻同期 [36]

カラオケのように楽曲の再生と同期して色が変わる歌詞を表示するシステムである。歌詞を理解する力を拡張してくれる。

例えば、普段は歌詞を意識して聴かない人も、再生と同期して色が変わる歌詞を見たり、興味のある歌詞の一部をクリックしてそこから聴いたりすることで、その歌詞に注目して聴き、歌詞が伝えるメッセージをよりの確に把握できる可能性がある。

- INTER: 音楽音響信号に対する楽器イコライザ [37], Drumix: ドラムパートのリアルタイム編集機能付き音楽再生インタフェース [38]

INTERは、楽曲中の複数の楽器パート(トラック)のそれぞれの音量を変化させることで、楽器単位のイコライジングをして音量バランスを変更(リミキシング)できる音楽再生インタフェースである。また、Drumixは、あたかも他のドラマーが違うドラムパターンで演奏しているかのように、楽曲を再生中にリアルタイムにドラムパートを加工できる音楽再生インタフェース

である。個々の楽器パートを聞き分けて理解する力を拡張してくれる。

例えば、普段はドラム音の音色やドラムパターンを意識して聴かない人も、Drumix を使ってそれらを加工することで、バスドラムとスネアドラムを聞き分け、ドラムパターンが楽曲の印象にいかにか大きな影響を与えているかを把握できるようになる可能性がある。こうしたカジュアルな二次創作体験が、創作支援にも連続的に繋がっていき、「一億総クリエイター時代」の実現に資すると我々は考えている。

これらは当初は、音楽を能動的に楽しむことができる「能動的音楽鑑賞インタフェース」(Active Music Listening Interfaces) [15], [31], [39] として研究開発された成果であるが、「音楽理解力拡張インタフェース」としても位置づけられる。このうち、SmartMusicKIOSK のインタフェースは、能動的音楽鑑賞サービス「Songle」(<http://songle.jp>) [40] の一機能として体験できる。

3.1.2 音楽理解力拡張インタフェースの考察

我々は、上記の事例で提供していた

- 楽曲が内在している情報(音楽的要素)の視覚化
- 音楽再生と同期した情報提示
- 音楽再生や音楽的要素を制御できるインタラクティブなインタフェース

の一連の機能が、音楽理解力拡張インタフェースを実現する上で大切だと考えている。「理解は『見る』ことで深まる」ので、可視化や情報提示は重要である。さらに、「理解は『制作・編集過程を疑似体験する』ことで深まる」ので、Drumix のように、既存の楽曲をインタラクティブに改変して能動的に鑑賞できるインタフェースは重要である。これにより、ユーザがカジュアルに楽曲の変化を楽しむ過程で、「クリエイターがなぜそのように楽曲を創作・設計したか」という意図に気付くことができる。

そして、上記で歌詞を聴くきっかけや、ドラムパートを聴くきっかけを与えたように、ある音楽的要素に気づかせるきっかけを与えて、ユーザが見過ごしがちな点をサポートするアプローチは効果的である。また、能動的でインタラクティブであることで没入感が得られ、より深くその楽曲と向かい合える状況を作り出すことも有効である。

3.1.3 音楽理解力向上インタフェースへ向けて

以上は3章の(b-1)「誰でも音楽家」を音楽鑑賞の観点から実現する「音楽を聴く力を一時的に拡張する支援」として位置付けられるが、その先に考えられるのは、(a)「トレーニング」に対応する「音楽を聴く力自体の向上・トレーニング」を可能にする研究である。外国語を聴く力を向上したければ語学学校や語学教材という手段があるが、音楽を聴く力を向上したくても、それに類する手段は整備されていない。音楽に関する学校や教材の多くは音楽家(音楽の作り手)がターゲットであり、ここでいうエンドユーザ(音楽

専門家だけでなく一般の人々が「何が似ているのか」「どれくらいありふれているのか」を知ろうと思えば自在に把握して活用できる技術基盤を構築

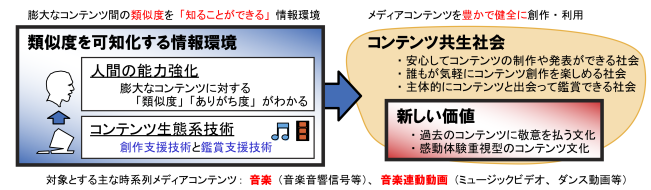


図1 OngaCREST プロジェクトの概要

の聴き手) が利用できるものは少ない。また、聴音トレーニングや絶対音感の習得も、「音楽を聴く力」を向上させて音楽をより深く理解するための手段とは違う。このような「音楽理解力向上インタフェース」(Music-Understanding Ability Training Interfaces) を実現することも、今後の重要な方向性の一つだと言える。料理の美味しさがわかる人が多いほど、美味しい料理は生まれやすくなるのと同様、音楽の魅力を理解できる人が多いほど、良い音楽は生まれやすくなる。そのための基礎としても、音楽理解力の拡張・向上は大切である。

3.2 OngaCREST プロジェクト: 類似度を可視化する情報環境

音楽において「類似度」(music similarity) の判断は本質的であり、その判断能力の拡張や活用に基づく様々な創作・鑑賞支援を目指した研究プロジェクト「コンテンツ共生社会のための類似度を可視化する情報環境の実現」(プロジェクト名: OngaCREST (音画 CREST/おんがくれすと) プロジェクト) を、(b) の「即効性のある支援」の例として紹介する。これは、科学技術振興機構(JST) 戦略的創造研究推進事業(CREST) 研究領域「共生社会に向けた人間調和型情報技術の構築」の平成23年度採択研究課題(研究代表者: 後藤 真孝, 研究分担者: 森島 繁生, 中村 聡史, 実施期間: 2011年10月~2017年3月) として推進中である。

本プロジェクトでは、図1に全体像を示したように、過去のコンテンツに敬意を払う文化を築きつつ、感動体験重視型の新たなコンテンツ文化が形成される社会を実現するために、膨大なコンテンツ間の類似度を人々が知ることができる(可視化する)情報環境を実現することを目的としている。そして、音楽の聴き方・創り方の未来を切り拓く技術開発により、音楽の楽しみ方がより能動的で豊かになる変化を日常生活に起こすことを目指している。本プロジェクトでは、「コンテンツ」として主に音楽あるいは音楽連動動画(ミュージックビデオ、ダンス動画等)を対象とする。

3.2.1 類似度に関する人間の能力の限界

人類の歴史の中で、今後もアクセス可能なコンテンツは単調増加し続けていく。そうすると、制作者にとっては自分の作品が埋もれやすくなり、視聴者にとっては選択がより難しくなっていく。さらに、似ているコンテンツも単調

増加して、本来は盗作ではないにも関わらず、盗作疑惑を招く事例が増えてしまう懸念がある。あらゆる創作は既存コンテンツの影響を受けており、無自覚に何らかの意味で部分的に類似してしまうのを避けることは難しい。

しかし、類似度に関する人間の能力には限界がある。目の前の二つを比較して類似度を判断するのは得意な一方、判断速度には限界があり、似たコンテンツを100万個の中から探すことはできない。さらに、人間は過去の経験に基づいて高度な判断ができる一方、ある事象が全体の中でどれぐらい確率的に起こりえるのか、という「ありがち度」の判断には限界がある。最近よく起きた事象を起きやすいと思いついたり、多数起きている事象でも遭遇しなければ滅多にないと誤解する。そのため、今後、誰もが創作と発表を楽しむ「一億総クリエイター時代」が到来しても、コンテンツが単調増加することによって、自分の作品が何かに似ていると糾弾されるリスクがあるために、人々が安心してコンテンツの制作や発表をしにくい社会になりかねない。

3.2.2 人々が類似度とありがち度を活用できる情報環境

そこで我々は、専門家だけでなく一般の人々が「何が似ているのか」「どれぐらいありふれているのか」を知ろうと思えば自在に把握して活用できる技術基盤を構築することが重要だと考え、本プロジェクトで取り組んでいる。それにより、人々が今後も安心してコンテンツの制作や発表を続けられるようにしたい。そして、「ありがち度」の高い事象（例えばコード進行やジャンルごとの慣例的な事象）は人類共有の知として活用できる創作支援技術を実現していくことで、誰もが気軽にコンテンツ創作を楽しめるようにする。さらに、類似度に基づいて新たなコンテンツと出会える鑑賞支援技術を実現していくことで、主体的にコンテンツと出会って鑑賞できるようにする。

それにより、論文のようにリファーされ再利用されたら喜びを感じられる新たな音楽文化が支援できれば、過去のコンテンツと共存共栄し、敬意を払う文化を築くことに貢献できる。コンテンツは「他にいかに類似していないか」に価値があるのではなく、新規性だけを追求しても人類は幸せになれないと我々は考えている。コンテンツは本来、人々をいかに感動させ幸せにするかに価値があり、感動させる魅力や完成度の高さがあれば価値があることを常識にしていきたい。むしろ学術論文等と同様に、多くの作品をリファーしてその土台の上に成り立っているからこそ価値が生まれる状況も望ましい。それにより、感動体験重視型のコンテンツ文化を築いていくことを目指している。

デジタルコンテンツ社会ではもはや忘却できず、ともすると、単調増加する過去の膨大なコンテンツに未来が押しつぶされかねない。本プロジェクトは、デジタル化による「忘却できない社会」において豊かで持続発展可能な「コンテンツ共生社会」を築く挑戦に位置づけられる。過去のコンテンツと未来のコンテンツとの共生を人々が実感でき

るようにすると共に、人間とコンテンツとの共生により膨大なコンテンツを楽しめる社会の実現を目指していく。

以上のように OngaCREST プロジェクトは、類似度に関する人間の能力、創作能力、鑑賞能力のいずれも、技術の力で拡張して支援する取り組みに位置づけられる。様々な中間成果の中で、上記の鑑賞支援技術に関連した成果の一部として、既に能動的音楽鑑賞サービス「Songle」(<http://songle.jp>) [40] と、多様な関係性に基づく音楽視聴支援サービス「Songrium」(<http://songrium.jp>) [41] を実現して誰でも利用可能にしてきた。

4. 計算機が創作して計算機が鑑賞する未来はあり得るのか

2章の「(2) 音楽家の能力の一部を全自動で可能にする研究（「仮想の音楽家」の実現）」の事例では、「自動創作」、「自動鑑賞」という言葉は取って出さなかったが、以下では、「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」というグランドチャレンジがあり得るのかについて、現段階ではやや浮世離れしていてSF的だと受け止められるかもしれないが、議論をしてみたい。

4.1 自動創作

まず、「自動創作」から考える。例えば、様々な未来の可能性の一つとして、もしかしたら、楽曲や歌詞すら技術の力によって全自動で生成され、それを人々が楽しむ文化も誕生するかも知れない。既に様々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、筆者自身も「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」という価値観が誰かによって打破されない保証はない。

既に、現在の動画共有サイト（「ニコニコ動画」「YouTube」等）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、音以外の証拠から掴むことは不可能になった。従来は、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、ポピュラー音楽の大半で用いられてきたが、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（とはいえ実際には、自動作曲や演奏、マスタリング等、歌声以外ですら人間同様に可能な技術はまだ存在していないので、そこから人間の介在はわかる）。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イトに投稿されても、気づけない環境で我々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、人間が気づかずに再生してコメントをつければ、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が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、より良い作品が生成可能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る状態を持ち、それを自己主張として歌で表現し始めたとしたら、我々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 (Consumer Generated Media, 消費者生成メディア) 作品 [42] 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、我々人間はどのように思うのだろうか。

4.2 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様々な音楽的要素や現象を記述したり [43]、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。動画共有サイト「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」[44] 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伴う形で「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が、もしそれが可能になったとしたら、我々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止めてくれたり、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、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可能性が高い。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つれない。

さらには、「自動創作」と「自動鑑賞」が共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る等の、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ある可能性の方が高いであろう。

5. 音楽情報処理がより一層世の中へ貢献していくために

今後も音楽情報処理の様々な技術の進展により、産業応用や社会への普及が進み、不可欠な技術として音楽情報処理が世の中に貢献できるのは間違いない。以下では、さらに人間にとって根源的な「音楽がもたらす感動とは」や「人類の幸せ・価値向上にどう寄与するか」という問題に対して考察を深めることで、仮説に基づく議論が中心にはなってしまうが、より一層大きな貢献をするための土台となる発想を広げていく。

5.1 音楽がもたらす感動とは

2 章の (2) 「音楽家の能力の一部を全自動で可能にする研究 (「仮想の音楽家」の実現)」で述べた自動〇〇や、4.1 章で述べた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。また、3 章の (b) 「即効性のある支援: 人間の能力を技術の力で一時的に向上させ支援するアプローチ」で能力が向上したわけではなく技術の支援で創作された音楽は、どう受け止められるのだろうか。それらが純粋に素晴らし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、少なくとも以下の三つの要因が絡み合っているという仮説を筆者は考えている。

(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持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、自動車で走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が、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、あの人がこうした理由で創作したコンテンツだから、感動する可能性がある。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。

本稿では上記の三つを取り上げるが、感動には他の要因もあり、例えば、他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）などが挙げられる。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時には、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、例えば動画共有サイト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づけて発表すれば、(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、音楽の創作や演奏等が手軽になって誰でも可能になったときに、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、あるキャラクタ（擬人化した仮想の身体性を持つ存在）[42]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。

人間が歌うだけで手軽に自然な歌声が合成できる我々の歌声合成技術「VocaListener（ぼかりす）」[45]でも類似した議論はあり、自然な歌声合成結果の楽曲があったときに、(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る一方で、それが手作業によるパラメータ調整でなく技術の力だと知っただけで、(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るのは興味深い（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、いざ理解が進めば(ii)の感動も広まる可能性はある）。また、我々の歌声合成文化の発展を真に願っている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」を感じて評価頂く場合もあり得る。

5.2 人類の幸せ・価値向上にどう寄与するか

音楽がもたらす感動について議論を進めてきたが、この音楽情報処理分野が長期的に目指すべきゴールの一つは、音楽情報処理の力で人類の幸せに寄与することである。「幸せ」とは心理的な現状認識の微分値（傾き）で決まり、絶対的な資源量では決まらないという仮説を筆者は考えている。これは観察と内省からきた仮説だが、限度はあるものの、資源量が少なくても、心理的な現状認識の微分値が正であれば、幸せを人間は感じやすい。一方、資源量がいくら多くても、心理的な現状認識の微分値が負であれば、不幸を人間は感じやすい。社会の「持続発展」が重要なのは、微分値を正にしやすいからで、「持続」だけでは微分値が0になる。

5.2.1 単位資源（リソース）辺りの価値向上

筆者は、解説「CGMの現在と未来」[42]で、

筆者は、今後は「単位エネルギー当たりの幸せ」（幸せ ÷ エネルギー）という新たな概念が重要になると提唱している。音楽等のメディアコンテンツは、豊かで人間らしい生活を送る上でもはや不可欠であり、幸せや充足感を得る手段の有無は生活の質に直結してくる。…特に音楽コンテンツは、必要なエネルギーが少なく繰り返しの視聴に耐える良質な娯楽であり、制作資源・エネルギーはデジタルコンテンツ制作環境の普及により下がってきた。N次創作は、既存の複数のコンテンツの良い点を凝縮して幸せを増すのに有効な手段として位置づけられる。結果的に、コンテンツがリサイクル（再利用）される側面もあるので、エネルギー効率の良いコンテンツ制作手段に意図しないうちになっていると解釈できる。こうした「単位エネルギー当たりの幸せ」を向上する（たとえば、消費エネルギーを減らしつつ幸せを増やす）ための研究開発は、今後ますます重要になっていくであろう。

と述べた。この「単位エネルギー当たりの幸せ」は、より汎用的な概念としては「単位資源（リソース）辺りの価値向上」と捉えることができる。古くに、単位資源として時間とコスト（人的資源を含む）のみを考え、価値向上として生産性向上を目指す方向性があったとすれば、現在の世の中では、単位資源として時間とコストに加えエネルギーを考え、幸せにつながる価値向上を目指すことが重要になっている。

5.2.2 音楽情報処理が寄与するには

情報化された社会においては、「心理的な現状認識」は、物理空間だけでなく情報空間にも大きく左右されるので、音楽情報処理（あるいは情報技術全般）にとって貢献できる余地が大きい。情報空間では、物理空間よりも資源の競

合が起きにくい状態にしやすく、そこでの価値向上を目指す技術は大きな可能性を秘めている。しかし、最適化や最大化は必ずしも真のゴールではなく、価値を最大化しようとすると、大きな資源が必要なことも多い。また、最大化が本当に達成されてしまったら、その定義上、その後は減少するしかないで、心理的な現状認識の微分値が負になってしまう。

よって、以上の考えがもし正しいのだとすれば、「いかに長期間、心理的な現状認識の微分値が正だと人間に感じさせ続けるか」が必要な挑戦となる。心理的な現状認識なので、実際の物理量と対応しておらず錯覚でもよい。音楽情報処理技術で人間を支援する場合にも、そうした観点で考察すると、従来とは違うアプローチが見えてくる可能性がある。例えば、3章(a)「トレーニング」では、徐々に能力が向上するから微分値が正だと長期間感じられるのであって、瞬時に世界最高の能力を獲得できる技術があったとしても、その後の微分値は正でなく、獲得できた本人が嬉しいとは限らない。(b)「即効性のある支援」では、一時的に能力を拡張していたのが実は重要で、支援を受けて能力向上を感じられる瞬間は微分値が正であり、使用をやめて再び使用することで、何度も正の微分値が得られる利点があった(ただし、使用をやめる瞬間には微分値が負となるのを、いかに認識させにくくするかも重要であることがわかる)。

さらに、コンテンツが単調増加し続ける中で、新たなコンテンツのクリエイターの「心理的な現状認識の微分値が正」な状態を達成する方法を考えてみる。すると、現在はコンテンツを発表する場が全国大会しかないのだとすれば、スポーツにおける全国大会と地区大会のような考え方が重要になる。例えば、クリエイターが、全国大会のように実力の違う人々とコンテンツの品質を競い合うとそれが埋もれてしまう(資源の競合が起きる)可能性が高くなるので、地区大会のように似た実力を持ったローカルな場で切磋琢磨し、埋もれずに鑑賞してもらえる(資源の競合が起きにくい)場が普及するとよいのではないだろうか。そこからステップアップしていく過程で、「心理的な現状認識の微分値が正」だと長期間感じさせられる可能性がある。

筆者は、以上のすべてが新規な考えだと主張しているのではなく、同様の仮説を考えている人たちは他にもいることを想定している。それでも、音楽情報処理が人類の幸せに寄与することを目指す重要性は変わらない。

6. おわりに

本稿では、音楽情報処理がより一層発展し、未来を切り拓いていくことを願って、敢えて刺激的な議論を試みた。具体的な事例としては、紙面の都合で、歌声合成や、音楽理解力拡張インタフェース、OngaCRESTプロジェクトを中心に議論を進めたが、音楽情報処理コミュニティには、

同じ文脈で議論できる研究事例は多数あり、今後も様々な研究者の活躍により、大きく発展していくことが期待される。特に、音楽情報処理が人類の幸せに寄与する貢献をしていくには、様々なコミュニティの研究者が協力し、議論を深めながら取り組んでいくことが不可欠である。学問のあり方が社会に開かれる形に変容しつつある現在、真に必要なのは、より広い視点から学問の再編を促すような分野の壁を越えた活動である。信号処理、機械学習、Web、インタラクション、音声、聴覚、音響、画像等の様々な分野とより密接に融合するような展開を期待したい。

謝辞 本研究の一部はJST CREST「CrestMuseプロジェクト」「OngaCRESTプロジェクト」の支援を受けた。本稿の議論には、これまで様々な機会に筆者と議論して下さった方々の知見も反映されており、日頃の学会やプロジェクト、グループ内での刺激的な議論に深く感謝する。

参考文献

- [1] 後藤真孝：研究会千夜一夜：音楽情報処理の研究を始めませんか？—音楽情報科学研究会(SIGMUS)—、情報処理(情報処理学会誌)、Vol. 48, No. 9, pp. 1046–1047 (2007).
- [2] Pardo, B.(ed.): *Special issue: Music information retrieval*, Communications of the ACM, Vol. 49, No. 8, pp. 28–58 (2006).
- [3] 片寄晴弘, 白木善尚(編): 特集「音楽情報処理技術の最前線」, 情報処理(情報処理学会誌), Vol. 50, No. 8, pp. 709–772 (2009).
- [4] 竹中 毅(編): 特集「音楽とOR」, オペレーションズ・リサーチ(システム制御情報学会誌), Vol. 54, No. 9, pp. 539–575 (2009).
- [5] 伊田政樹(編): 「音楽制作と情報処理の友好関係」特集号, システム/制御/情報(システム制御情報学会誌), Vol. 56, No. 5, pp. 199–262 (2012).
- [6] 後藤真孝, 奥乃 博: 特集「CGMの現在と未来: 初音ミク, ニコニコ動画, ビアプロの切り拓いた世界」, 情報処理(情報処理学会誌), Vol. 53, No. 5, pp. 464–494 (2012).
- [7] 後藤真孝, 平田圭二: 解説“音楽情報処理の最近の研究”, 日本音響学会誌, Vol. 60, No. 11, pp. 675–681 (2004).
- [8] Downie, J. S.: The music information retrieval evaluation exchange (2005–2007): A window into music information retrieval research, *Acoustical Science and Technology (edited by the Acoustical Society of Japan)*, Vol. 29, No. 4, pp. 247–255 (2008).
- [9] 後藤真孝, 齋藤 毅, 中野倫靖, 藤原弘将: 解説“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).
- [10] 嵯峨山茂樹, 亀岡弘和: 解説“自動採譜技術の展望”, 日本音響学会誌, Vol. 64, No. 12, pp. 715–720 (2008).
- [11] 帆足啓一郎, 上月勝博, 菅谷史昭: 楽曲配信サービスを支える音楽情報検索技術, 電子情報通信学会誌, Vol. 88, No. 7, pp. 529–534 (2005).
- [12] 片寄晴弘: 音楽生成とAI, 人工知能学会誌, Vol. 19, No. 1, pp. 21–28 (2004).
- [13] 北原鉄朗: 私のブックマーク「音楽情報処理」, 人工知能学会誌, Vol. 24, No. 6, pp. 921–929 (2009).
- [14] 後藤真孝, 緒方 淳: 音楽・音声の音響信号の認識・理解研究の動向, コンピュータソフトウェア(日本ソフトウェア科学会論文誌), Vol. 26, No. 1, pp. 4–24 (2009).

- [15] 後藤真孝: 解説 “音楽鑑賞・音声入力をより豊かに～能動的音楽鑑賞インタフェースと音声補完シリーズ～”, 日本バーチャルリアリティ学会誌, Vol. 13, No. 3, pp. 140-144 (2008).
- [16] 片寄晴弘: 音楽とエンタテインメント, 日本バーチャルリアリティ学会誌, Vol. 9, No. 1, pp. 20-24 (2004).
- [17] 帆足啓一郎: 音楽情報の検索, 映像情報メディア学会誌, Vol. 64, No. 5, pp. 701-707 (2010).
- [18] 後藤真孝: 解説 “「初音ミク」はなぜ注目されているのか”, 電気学会誌, Vol. 132, No. 9, pp. 630-633 (2012).
- [19] 後藤真孝: 音楽情報学, 情報処理 (情報処理学会誌), Vol. 51, No. 6, pp. 661-668 (2010).
- [20] 剣持秀紀: 歌声合成技術と VOCALOID, ヒューマンインタフェース学会誌, Vol. 10, No. 2, pp. 95-98 (2008).
- [21] 剣持秀紀: 解説 “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, Vol. 67, No. 1, pp. 46-50 (2011).
- [22] Katayose, H., Hashida, M., Poli, G. D. and Hirata, K.: On Evaluating Systems for Generating Expressive Music Performance: the Rencon Experience, *Journal of New Music Research*, Vol. 41, No. 4, pp. 299-310 (2012).
- [23] 嵯峨山茂樹, 酒向慎司, 堀 玄, 深山 寛: 確率的手法による歌唱曲の自動作曲, システム/制御/情報 (システム制御情報学会誌), Vol. 56, No. 5, pp. 219-225 (2012).
- [24] Fukayama, S., Saito, D. and Sagayama, S.: Assistance for novice users on creating songs for Japanese lyrics, *Proc. of ICMC 2012* (2012).
- [25] 中野倫靖, 後藤真孝, 平賀 譲: MiruSinger: 歌を「歌って/聴いて/描いて」見る歌唱力向上支援インタフェース, 情報処理学会インタラクティブ 2007 論文集 (インタラクティブ発表), pp. 195-196 (2007).
- [26] 平井重行, 片寄晴弘, 井口征士: 歌の調子外れに対する治療支援システム, 信学論 (D-II), Vol. J84-D-II, No. 9, pp. 1933-1941 (2001).
- [27] 竹川佳成, 寺田 努, 塚本昌彦: 運指認識技術を活用したピアノ演奏学習支援システムの構築, 情報処理学会論文誌, Vol. 52, No. 2, pp. 917-927 (2011).
- [28] 大島千佳, 西本一志, 鈴木雅実: 家庭における子どもの練習意欲を高めるピアノ連弾支援システムの提案, 情報処理学会論文誌, Vol. 46, No. 1, pp. 157-171 (2005).
- [29] Baba, T., Hashida, M. and Katayose, H.: “VirtualPhilharmony”: A Conducting System with Heuristics of Conducting an Orchestra, *Proc. of NIME 2010*, pp. 263-270 (2010).
- [30] 中野倫靖, 後藤真孝: VocaRefiner: 歌を歌って歌い直して統合できる新しい歌声生成インタフェース, WISS 2012 論文集 (2012).
- [31] 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, 情報研報音楽情報科学 2007-MUS-70-9, pp. 59-66 (2007).
- [32] Goto, M.: Augmented Music-Understanding Interfaces, *Proc. of SMC 2009 (Inspirational Session)* (2009).
- [33] Goto, M.: Music Listening in the Future: Augmented Music-Understanding Interfaces and Crowd Music Listening, *Proc. of the AES 42nd International Conf. on Semantic Audio*, pp. 21-30 (2011).
- [34] 後藤真孝: SmartMusicKIOSK: サビ出し機能付き音楽試聴機, 情報処理学会論文誌, Vol. 44, No. 11, pp. 2737-2747 (2003).
- [35] Goto, M.: A Chorus-Section Detection Method for Musical Audio Signals and Its Application to a Music Listening Station, *IEEE Trans. on ASLP*, Vol. 14, No. 5, pp. 1783-1794 (2006).
- [36] Fujihara, H., Goto, M., Ogata, J. and Okuno, H. G.: LyricSynchronizer: Automatic Synchronization System Between Musical Audio Signals and Lyrics, *IEEE Journal of Selected Topics in Signal Processing*, Vol. 5, No. 6, pp. 1252-1261 (2011).
- [37] Itoyama, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H. G.: Instrument Equalizer for Query-by-Example Retrieval: Improving Sound Source Separation Based on Integrated Harmonic and Inharmonic Models, *Proc. of ISMIR 2008*, pp. 133-138 (2008).
- [38] Yoshii, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H. G.: Drumix: An Audio Player with Functions of Realtime Drum-Part Rearrangement for Active Music Listening, 情報処理学会論文誌, Vol. 48, No. 3 (2007).
- [39] Goto, M.: Active Music Listening Interfaces Based on Signal Processing, *Proc. of ICASSP 2007* (2007).
- [40] 後藤真孝, 吉井和佳, 藤原弘将, Mauch, M., 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, Vol. 54, No. 4, pp. 1363-1372 (2013).
- [41] 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).
- [42] 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), Vol. 53, No. 5, pp. 466-471 (2012).
- [43] Goto, M.: A Real-time Music Scene Description System: Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-world Audio Signals, *Speech Communication*, Vol. 43, No. 4, pp. 311-329 (2004).
- [44] Yoshii, K. and Goto, M.: MusicCommentator: Generating Comments Synchronized with Musical Audio Signals by a Joint Probabilistic Model of Acoustic and Textual Features, *Proc. of ICEC 2009*, pp. 85-97 (2009).
- [45] 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, Vol. 52, No. 12, pp. 3853-3867 (2011).