

# 音楽情報処理が切り拓く 音楽体験の未来

後藤真孝

ことう まさたか

産業技術総合研究所 情報・人間工学領域(音楽情報処理)

音楽は技術と強く結びついており、新しい技術の登場によって、音楽の表現や鑑賞方法は大きく変わってきた。伝統的なピアノやギターも、発明当時は最先端技術にもとづく楽器だった。シンセサイザーや歌声合成の普及は、新たな音楽文化を生み出し、音楽をインターネットで聴くことも当たり前になった。こうした発展を支える研究分野が「音楽情報処理」である。では、最近は何が起きていて、音楽体験の未来はどうなっていくのだろうか。

## 技術が新たな音楽体験・表現を生み出す

技術の発展によって、さまざまな新たな音楽体験や音楽表現が生み出されてきた。

今では伝統的な楽器と見なされるピアノでさえ、18世紀に発明された当時は革新的な楽器だった。ピアノの登場によって、音楽家は、より大きな音や広い音域、豊かな音色を、音楽表現として利用できるようになった。エレキギターやサクソフォンのようなさまざまな楽器も、技術の力で誕生した。

さらに技術が発展すると、自在に多様な音を作り出せるシンセサイザー(楽音合成)が発明され、ポピュラー音楽では欠かせない楽器となった。コンピュータの普及に伴って音楽制作環境のデジタル化も進み、2007年以降は、歌声合成ソフトウェア「初音ミク」によって歌声合成が注目を集め、新たな表現が生み出されてきた。

音楽を聴いて楽しむ体験も、技術の発展とともに大きく変化した。かつては生演奏を近くで聴く

しかなかったが、たとえばピアノの発明によってより大きな音が出せるようになると、それを用いた演奏会が、聴衆にとって新たな音楽体験となった。19世紀にレコードが発明されて、音の録音・再生が可能になり、さらにラジオが発明されて音の遠隔伝送も可能になると、物理的にその場になくとも音楽を聴けることが当たり前になった。記録媒体も、アナログのレコードやカセットテープから、デジタルのコンパクトディスク(CD)に発展して、音楽体験を大きく変えた。

インターネットの普及によって音楽配信が可能になると、当初はダウンロードして手元に保存する形だったが、やがてはオンラインで常時アクセスする形へと移行し、音楽を聴く環境はさらに大きく変化した。音楽ストリーミング(定額制音楽配信)サービスや動画共有サービスで音楽を鑑賞するのが一般化し、好きな音楽を好きなときに好きなだけ楽しめる、夢のような時代が到来した。誰かが自宅のパソコンで作った曲が、すぐに世界中で聴けるようになったのも画期的だった。

これらはすべて、多くの方々が新たな音楽体験・表現を生み出そうと貢献してきたおかげである。特に過去30年間の変化は、コンピュータを中心とした技術の発展によるところが大きい。

では、最近は何が起きていて、これからどうなっていくのだろうか。それを議論し、研究開発する研究分野が、「音楽情報処理」である。既に確立した研究分野となって久しいので、聞いたことがある人も多いかもしれない。音楽情報処理とは、コンピュータ上で音楽のあらゆる側面を扱う研究

分野であり、実はコンピュータの発明とほぼ同時期から、音楽への応用が始まっていた。特に音楽がデジタル化されて以降、その発展は加速し、もはや音楽を楽しむのにスマートフォンやパソコンを使うことが当たり前になるぐらい、音楽情報処理は、学術・産業・社会・文化的に重要性を増している。

筆者はこの研究分野で30年以上にわたり、「音楽理解技術」とその応用を研究開発してきた(図4(a))。

## 音楽の中身を自動解析するコンピュータ

音楽理解技術とは、ある楽曲の音響信号(耳で聴くような音波)をコンピュータに入力し、メロディやビート(拍)、サビ(楽曲中で盛り上がる代表的な区間)、コード進行(ハーモニー)のような音楽の中身を自動解析して出力する技術である。人間は音楽を聴くとそうした中身を理解できるが、コンピュータでその理解を実現することは、当初はできなかった。その理由は、音楽が、複数の楽器音や歌声が混ざり合った、きわめて複雑な音響信号だからである。音の一つなら解析は容易でも、ポピュラー音楽のように数十個の音が鳴っていると難しい。近年でも、ピアノの独奏を入力して楽譜を出力する自動採譜の精度は上がったが、ポピュラー音楽に対する自動採譜は、まだ実現されていない。

ただ、人間は音楽を聴いて楽譜を頭に思い浮かべたりしなくとも、メロディを口ずさんだり、ビートに合わせて手拍子を打ったり、サビがわかったりしている。同様にコンピュータでも、すべての音を楽譜化しようとはせずに、そうした音楽の中身を直接推定する音楽理解技術が、1990年代から発展してきた。

たとえば、ポピュラー音楽のサビは、楽曲によって音響的な特徴が異なり、必ずしも音量が大きかったり音が高かったりするわけではないが、一番聴かせたい盛り上がる部分なのでたくさん繰り返される。筆者はこのことに着目して、楽曲中の繰り返し構造を解析し、多く繰り返される区間で、

かつ、サビらしい繰り返し方がされている区間(「Aメロ→Bメロ→サビ」のような長いまとまりが繰り返されていたらその最後が「サビらしい」など)を求めることで、サビの自動検出を実現した<sup>2</sup>(実際にはもっと複雑な信号処理だが、ここでは単純化して説明している)。

## 好きなところを聴きたいように聴ける技術

この音楽理解技術は、さまざまなアプリケーションの土台となる重要な基盤技術である。

たとえば、サビが自動検出できると、試聴して好みの楽曲か判断したいときに、先頭から聴くだけでなく、いきなりサビから聴けて便利になる。これを世界で初めて可能にしたのが、サビ出し機能付き音楽試聴機「SmartMusicKIOSK」<sup>2</sup>である。サビの位置を含む繰り返し構造が自動解析され、それが図1のように可視化されていて、「サビ出しボタン」を押すと、次のサビの先頭へジャンプして聴くことができる。これにより、いろいろな楽曲のサビだけを次々と試聴して、好みの楽曲を見つけやすくなった。これは、従来の楽曲単位での操作体系に対して、楽曲内部の区間単位での操作体系が初めて追加された音楽鑑賞インタフェースという点で画期的だった。

実は、人類が興味のない楽曲をボタンひと押しで瞬時に飛ばすことが初めて可能になったのは、音楽がデジタル化され、CDプレーヤーが登場したときだった。SmartMusicKIOSKは、楽曲内で、

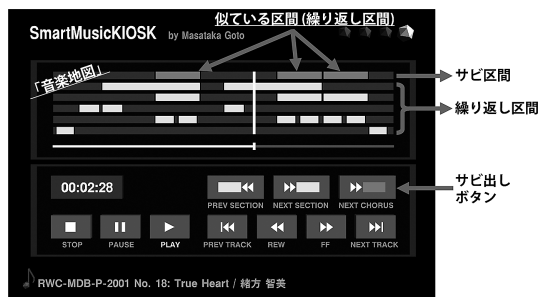


図1—サビ出し機能付き音楽試聴機「SmartMusicKIOSK」この楽曲では「サビ区間」が3回繰り返されていることが可視化されていて、「サビ出しボタン(NEXT CHORUS)」を押すと、次のサビの先頭へジャンプ(瞬時に早送り)して聴くことができる。図4(b)にアクセスすると、実際の動作の様子を見ることができる。



図2—音楽チェックサービス「Kiite Check」

歌声合成技術による歌唱を含む約58万曲の音楽動画の中から、自分に対してオススメされた10曲のサビだけを次々と試聴(チェック)することで、お気に入りの楽曲を発掘することができる。各楽曲は自動検出されたサビから再生され、5秒以上聴くと、ボタン操作や下から上へのスワイプ操作によって、次の楽曲に進むことができる。試聴した楽曲が好みだったら自分の「お気に入り曲」として登録しておき、あとで楽曲の先頭から最後までじっくり聴くこともできる。図4(c)にアクセスすると、無料で利用できる。

興味のない区間をボタンひと押しで瞬時に飛ばすことを初めて可能にした。こうして原曲の時系列に沿わずに、「好きなところを聴きたいように聴ける」ようになったメリットは大きく、試聴以外にも有用である。たとえば、繰り返されるサビだけを聴き比べやすいので、同じメロディでも歌詞やアレンジが変化していく様子を、よりの確に把握できる。可視化された繰り返し構造は、楽曲構造を意識して聴くきっかけにもなる。

このサビ区間検出技術は、既に実用化されている。図2の誰でも無料で使える音楽チェックサービス「Kiite Check」では、毎回変わるオススメの10曲のサビを次々と試聴して、お気に入りの楽曲を発掘できる。スマートフォンで使える音楽アプリ「ボカコレ」では、「サビメドレー」機能が人気で、ランキングやお好みのプレイリストを音楽紹介番組のようにサビだけメドレー形式で再生できる。

これらは、楽曲をそのまま再生して聴くだけでなく、聴き手が音楽再生に能動的に介入しやすいインタフェースである点が特徴的である。

## 鑑賞をもっと能動的に

このように音楽理解技術は、人々の音楽の聴き方を、より能動的で、便利で、豊かなものにする力をもっている。そこで、従来の受動的な鑑賞とは違う、能動的な音楽鑑賞を可能にする「能動的音楽鑑賞インタフェース」<sup>3)</sup>というコンセプトを

筆者は提唱した。ここでの「能動的」という言葉は、音楽の創作を意味するわけではなく、音楽鑑賞を楽しむ上でのあらゆる能動的なインタラクションを意味している。

実は昔から、音楽の鑑賞は必ずしも受動的ではなかった。たとえば生演奏を聴いて、拍手や手拍子、歓声などで能動的に反応することはよくある。音楽のプレイリストを作成して再生曲順を自分で決めたり、音質調整つまみで自分好みの音質(周波数特性)に調整したりするのも、能動的な楽しみ方である。

能動的音楽鑑賞インタフェースは、さらに多様で新しい楽しみ方を、音楽理解技術によって可能にする。具体的には、少なくとも、音楽の(1)再生、(2)加工、(3)検索・ブラウジングの三つの異なる方向性が考えられる<sup>3)</sup>。

(1)の音楽再生の先駆的な代表例は、さきほど紹介したSmartMusicKIOSKである。他にも、歌詞にもとづいてジャンプできる能動的音楽鑑賞インタフェースなど、いろいろと研究開発されてきた。

(2)の音楽加工の先駆的な代表例に、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「Drumix」<sup>4)</sup>がある。ユーザは、既存の楽曲を再生中に、ドラム音の音量や音色、ドラムパターンを自分好みに変更して、雰囲気を変えて楽しめる。既存の楽曲をそのまま聴くだけでなく、より自分好みになるように能動的に加工する方法は、他にもさまざまな研究者が取り組んできた。

(3)の音楽検索・ブラウジングについては、膨大な未知の楽曲の中で、良い楽曲を埋もれさせずに発掘する重要性が増している。そのためには、曲名やアーティスト名の一覧から選んだり、テキスト検索で選んだりする古典的な方法では不十分で、新たな方法で音楽を検索、推薦、発掘できるさまざまなインタフェースが研究開発されてきた。たとえば音楽発掘サービス「Kiite」(図4(d))では、好きな楽曲に雰囲気が近い(曲調が似た)楽曲を検索できる機能、お気に入り楽曲の中身(曲調)を考慮した音楽推薦を受けられる機能などが実現されている。特に、ユーザ自身が複数の推薦エンジンを能動的に使い分けられる機能は先進的である<sup>5</sup>。さきほど紹介したKiite Checkでは、未知の楽曲のサビを効率よく試聴して発掘できる(図4(c))。

## 初音ミクが変えた歴史

ここまで鑑賞側の話をしてきたが、創作側でも技術の発展が大きな変化をもたらした。特に、冒頭で言及した歌声合成ソフトウェア「初音ミク」(2007年8月31日発売)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を打破した。これにより「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。音楽文化の歴史を変えた快挙だった。

「初音ミク」は、ヤマハ株式会社の歌声合成技術 VOCALOID を用いて、あたかも仮想の歌手が歌っているかのような「歌声」を合成するソフトウェア製品として、クリプトン・フューチャー・メディア株式会社が開発し発売した。その開発では、仮想の歌手をイメージしやすいように「初音ミク」というキャラクターのイラストを製品パッケージに描き、その歌声データベースを高品質に収録したことが本質的な貢献だった。キャラクターという身体性をもたせたことで初めて、「誰が歌っているのか」、「誰に歌わせたいのか」が明確になった。それによって、「初音ミク側の視点からの歌詞をつけた曲」のような、彼女にしか歌えない曲も発表され、「彼女のために曲を作る文

化」が誕生したのである。

歌声合成の技術自体は1980~1990年代にもいろいろと登場していた。しかし、まさか合成された歌声がメインボーカルになるとは誰も思っていなかったのである。初音ミクによって「このキャラクターに歌わせたいから曲を作る」という強い動機が生まれ、「このキャラクターの歌だから聴きたい」という積極的な楽しみ方が生まれたことは、筆者も含めた当時の研究者にとって完全に予想外だった。これによって、人間の代わりに歌声合成を用いるのではなく、歌声合成でなければならない表現を含む音楽文化が誕生した<sup>6</sup>。

## 「自動生成」時代の感動と創作のゆくえ

当時はこれを一過性のブームと勘違いする人がいたので、筆者は2010年に「歌声合成技術がいつの日か普及することは歴史的必然である」と主張した。実際にその後、歌声合成は、あって当たり前前の技術として普及している。

さらに筆者は、2012~2013年には、「人間の歌声でなければ聴く価値がない」という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<sup>8</sup>と主張した。当時は、楽曲の音響信号を1曲丸ごと自動生成する技術は存在せず、実現可能性はまだ低い状態ではあった。しかし、技術が進歩したときの未来予想として、「自動生成された音楽は、それ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か」を議論したのである。そして、コンテンツ全般(ここでは音楽)がもたらす感動には、(i)コンテンツ自体の感動、(ii)オリエンティック的感動、(iii)文脈的感動の三つの要因が少なくとも絡み合っているという仮説を、筆者は提示した<sup>8</sup>。

この(i)は、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。(ii)は、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。たとえば、まったく同じ音楽でも、人間が目の前で早弾きのすごい演奏をしていれば感動できるが、それが自動演奏されていたら感動できないことに相当する。これは、オリンピックで選手が走る横を、自動車で走って先にゴールしても感動しないのと同様であり、自分と同じ人間がその行為をしている事実が感動を生んでいるということである。最後の(iii)は、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。たとえば、同じ楽曲でも、それが誕生した社会・文化・個人的背景を知っていることで感動することがある。この状況でこういう表現をした、あの影響を受けて生まれた、あの人がこの理由で創作した、という文脈に感動する。実は(ii)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。感動には他の要因もあり、コンテンツ側の文脈でなく鑑賞者側の文脈として、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。

現在では、音楽情報処理や機械学習の進歩により、音楽の自動生成はある程度可能になっている。それを人々が受け入れるかどうかは、この(i)～(iii)の感動の要因でいろいろと考察できる。ただし、歌声合成のときは違って、「楽曲生成(音楽生成AI)でなければならぬ表現を含む音楽文化が誕生した」という状況には至っていない点は、注目に値する。現在の楽曲生成は、まだ未成熟な要素技術の状態なのである。

筆者は、人間のクリエイターを応援する立場で研究開発をしている。そのため、こうした楽曲生成技術が将来、創作支援技術として発展したときに、どのようにクリエイターの幸せにつながり、人々の音楽体験や音楽表現を豊かにできるかに関心がある。2012年に筆者は、「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だろうか。筆者はそうは思わない。なぜなら、我々人類は歌うことをやめないからである。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。また

聴き手も、この人(自分の子供、友だち、好きな歌手)が歌うから、と思って聴く。そこに初音ミクという選択肢が増えるに過ぎないのである」と主張した<sup>6</sup>。同様に、楽曲生成技術が発展しても、人間のクリエイターは活躍するし、人類は創作することをやめないと、筆者は未来予想している。

## 未来をみんなと一緒に切り拓く

このように、音楽情報処理は、鑑賞と創作の両面で、音楽体験の未来を切り拓き続けている。その研究成果の多くは、今では当たり前の技術として広く利用されている。現在もさまざまな研究者が、この先の未来に向けた研究開発に取り組んでいるが、筆者らの最新の研究成果を紹介して結びとしたい。

筆者は、この先の未来において、「コピー不可能な能動的体験」(図4(a))にもとづく新たな価値を創出することが重要になると考えている。もちろん、音楽を聴いて感動すること自体の重要さは今後も変わらないが、誰でもアクセスできるコンテンツを聴くような従来型の「コピー可能な受動的体験」だけでなく、コンテンツの能動的な鑑賞・創作体験を可能にすることが新たな価値を生み出す。そのためには、今しか味わえないパーソナルな体験をいかに創出するか、単なる自動生成とは違う能動的体験やクリエイティブな体験をいかに生むかが鍵となる。

たとえば、図3の音楽発掘カフェ「Kiite Cafe」では、みんなと一緒に同じ楽曲を同じ瞬間に聴きながら、リアルタイムにコミュニケーションできる<sup>9</sup>。各自が「イチ推しリスト」としてみんなに聴いてもらいたい楽曲を持ち寄り、そこから選ばれた楽曲と一緒に聴いているユーザが気に入って「いいね(ハートマーク)」を押すと、そのユーザのアイコンの右上に、ハートのアニメーションが出る。そこで、自分の好きな楽曲が流れたときに、他の人たちが気に入ってハートが表示されると嬉しくて感激する。これはとても衝撃的で、「自分の好きな楽曲を他の人たちが気に入るまさにその瞬

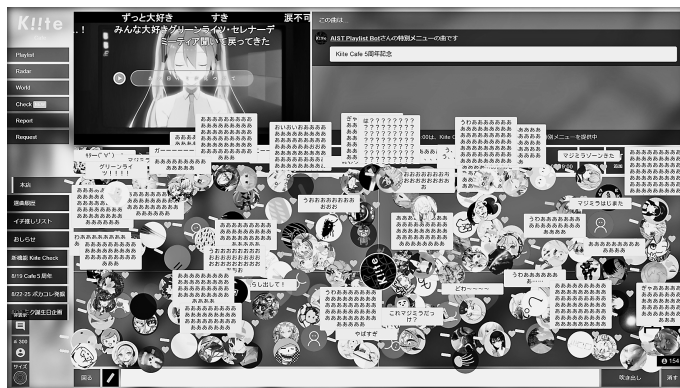


図3—音楽発掘カフェ「Kiite Cafe」

集まっているユーザたちの好きな楽曲(音楽動画)がランダムに再生され、それをみんなで同時に視聴できる。お気に入り(いいね)ボタンを押すとハートマークが出て、コメントを書くことができ、吹き出しが出る。ユーザは「リスナー」の役割に加えて、好みの楽曲を埋もれさせずに発掘して他の人にお勧めする「キュレーター」というクリエイティブな役割も果たしている。図4(e)にアクセスすると、実際の動作の様子を見ることができる。

間」を見ることができる感動は、ぜひ体験してもらいたい。さらに、年に10回以上、ペンライトを振りながらライブのように一緒に聴けるイベントも開催されている。楽曲自体は動画共有サービスでいつでも無料で聴けるのに、その場に集まってみんなで聴く体験自体は「コピー不可能な能動的体験」となり、人はそこに特別な価値を感じるのである。

他にも、音楽は映像表現を伴ったミュージックビデオの形で楽しめることが一般的になったが、「ビデオの次の音楽体験はアプリ」という新概念を、筆者らは世界に先駆けて提唱している。近年、歌詞が音楽に合わせて動くミュージックビデオは「リリックビデオ」とよばれて人気を集めているが、単なる動画ファイルなので、誰がいつ見ても同じである。それに対して、筆者らは2020年に、インタラクティブで体験するたびに变化する「リリックアプリ」<sup>10</sup>という新たなコンテンツ表現を提案した。さらに、そのインタラクティブな歌詞駆動型視覚表現を開発しやすくするソフトウェア技術「TextAlive App API」<sup>10</sup>を公開した結果、それをういたプログラミング・コンテストが2020年から毎年開催されている。6年間で318作品の応募が国内外からあって、その中から入選した

(a)産総研マガジン「初音ミクも研究対象！未来の音楽体験を作る研究」(多数のデモンストレーション動画)  
URL: [https://www.aist.go.jp/aist\\_j/magazine/20240221.html](https://www.aist.go.jp/aist_j/magazine/20240221.html)



(b)SmartMusicKIOSKの動作を動画で見る  
URL: <https://youtu.be/Cjh9IsDAE14&t=2321>



(c)Kiite Check を利用する  
URL: <https://check.kiite.jp>



(d)Kiiteの動作を動画で見る  
URL: <https://youtu.be/1V2HRxKQ1fs&t=1107>



(e)Kiite Cafeの動作を動画で見る  
URL: [https://youtu.be/4SNjMz0\\_5ko&t=143](https://youtu.be/4SNjMz0_5ko&t=143)



(f)リリックアプリを体験する  
URL: <https://textalive.jp/apps>



図4—デモンストレーション動画や参考情報にアクセスするためのQRコードとURL

60 作品が無料公開されている。音楽体験の未来を感じさせる多くのアイデアが、リリックアプリとして誰でも体験できる状態になっているので、図4(f)に掲載したQRコードあるいはURLでぜひ体験してもらいたい。これも楽曲自体はいつでも聴けるものだが、リリックアプリを実行している本人しか味わえないパーソナルな体験は、「コピー不可能な能動的体験」で、特別な価値を感じる。

これからも、音楽体験・音楽表現の未来は、人々の力と技術の力で切り拓かれていく。ここに紹介しきれなかった魅力的な事例はあまりに多く、これを読んで興味をもった方々には、ぜひ、多くの研究者や企業の実現されてきたそうした事例にも目を向けて欲しい。これをきっかけに音楽情報処理に興味をもって、研究開発者や利用者、応援者などのさまざまな立場で関わる人が一人でも増え、一緒に「未来の当たり前を創る」活動を発展させていければと願っている。

#### 文献

1—後藤真孝・他:『音楽情報処理』。コロナ社(2023)

- 2—M. Goto: IEEE Transactions on Audio, Speech, and Language Processing, **14**, 1783(2006)
- 3—後藤真孝: 情報処理学会 音楽情報科学研究会 研究報告, **2007-MUS-70-9**, 59(2007)
- 4—K. Yoshii et al.: IPSJ Journal, **48**, 1229(2007)
- 5—K. Tsukuda et al.: Multimedia Tools and Applications, **84**, 46973(2025)
- 6—後藤真孝: 情報処理学会誌, **53**, 466(2012)
- 7—後藤真孝: 情報処理学会誌, **51**, 661(2010)
- 8—後藤真孝: 情報処理学会 音楽情報科学研究会 研究報告, **2013-MUS-99**, 1(2013)
- 9—K. Tsukuda et al.: IEICE Transactions on Information and Systems, **E106-D**, 1906(2023)
- 10—加藤淳・後藤真孝: 情報処理学会論文誌, **66**, 273(2025)

#### 後藤真孝 ごとう まさたか

国立研究開発法人産業技術総合研究所 上級首席研究員。1998年早稲田大学大学院 理工学研究科 博士後期課程修了。博士(工学)。2024年より産業技術総合研究所で初の上級首席研究員。2009～2017年にIPA未踏IT人材発掘・育成事業PM, 2016～2022年にJST ACT-I研究総括を兼任。現在, JST創発PO, 早稲田大学 客員教授, 筑波大学連携大学院 教授などを兼任。日本学士院学術奨励賞, 日本学術振興会賞, ドコモ・モバイル・サイエンス賞, 市村学術賞, 星雲賞など, 77件受賞。未来のメディアコンテンツ体験をより豊かにする技術の確立を目指し, 音楽理解技術とそれにもとづくインタフェース・サービスを中心とした音楽情報処理等の研究を34年間推進。編著に『音楽情報処理』(コロナ社, 2023年)。

URL: <https://staff.aist.go.jp/m.goto/>