

特集 ■ インタラクシオン技術の新展開

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～



後藤真孝

産業技術総合研究所

GOTO MASATAKA

1. はじめに

音楽や音声の音響信号を自動的に認識・理解できる技術は日々進歩しているが、エンドユーザは、まだその恩恵を充分には享受していない。そのためには、日常的な場面で効果的に使える音楽インタフェース、音声インタフェースの研究が重要となる。本解説では、そうした音響信号の自動理解技術が実際にエンドユーザの役に立ち、重要であることを示すために我々が取り組んでいる二つの研究アプローチ「能動的音楽鑑賞インタフェース」[1]と「音声補完シリーズ」[2]を紹介する。

「能動的音楽鑑賞インタフェース」は、「音楽音響信号理解研究が、音楽の聴き方をどのように豊かにできるか」を明らかにする取り組みである。例えば、楽曲構造を見ながら楽曲内で興味のない区間を簡単に飛ばしたり、楽曲中の楽器の音色を変更したり、音楽コレクションをブラウジングして興味のある楽曲やアーティストに出会ったりするためのインタフェース等を実現してきた。

一方、「音声補完シリーズ」は、「ユーザに意図的に非言語情報を発声してもらうことで、音声の持つ潜在能力をどのように引き出せるか」を明らかにする取り組みである。例えば、言い淀むと助けてくれる音声インタフェース機能や、不自然な非言語情報によって人間同士の会話中に音声認識が利用できる音声インタフェース機能等を実現してきた。

いずれも音楽鑑賞や音声入力をより豊かにするインタラクシオン技術であり、本解説では、こうした各事例の紹介を通じて、その背後にある「音楽・音声インタラクシオン」に関する我々の考え方を議論していく。

2. 能動的音楽鑑賞インタフェース

我々は 2003 年に、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エンドユーザ向け音楽インタフェースを実現する研究アプローチを提唱した[3]。そして、これを「能動的音楽鑑賞インタフェース」と名付けて、音楽音響信号理解に基づく様々な事例を実現する研究を進めている[1]。ここでの「能動的」という言葉は、音楽の創作は意味せず、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オン(音楽再生中の新たな支援や加工、楽曲の検索やブラウジング等)を意味する。「音楽との能動的なインタラクシオン」というと、通常は、音楽家だけが可能な楽しみだと考えられがちである。しかし、音楽家でない普通の人達が音楽を聴く行為は、必ずしも受動的とは限らない。例えば、生演奏を聴きながら、手拍子や歓声等によって能動的に働きかけたり、コンパクトディスク(CD)やプレイリスト上での楽曲の再生順を自分で決めたり、音質調整つまみで周波数特性を調整したりすることは、一種の能動的なインタラクシオンと言える。本研究では、それをさらに発展させ、既存の楽曲のモノラル音響信号を入力として、音楽的要素(サビ区間やメロディー、ドラム音等)を自動的に理解する技術によって、より能動的で豊かな音楽の聴き方を可能にすることを目指している。以下、具体的な事例を文献[1]から抜粋して紹介する。

2.1 SmartMusicKIOSK: サビ出し機能付き音楽試聴機

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。図 1 下側の「サビ出しボタン(NEXT CHORUS)」を押すことで、次のサビの先頭へジャンプ(瞬

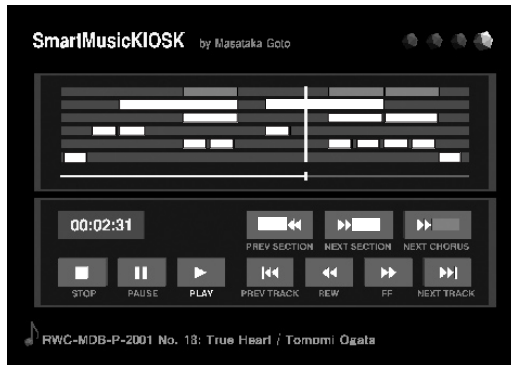


図 1 SmartMusicKIOSK

視覚化された楽曲の繰り返し構造（上側の「音楽地図」ウィンドウ）を見ながら、楽曲の様々な箇所を自在に行き来して能動的に聞くことができる

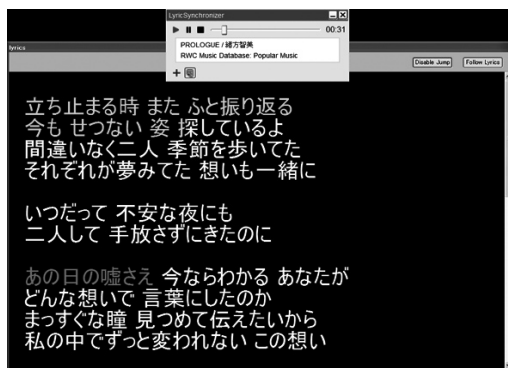


図 2 LyricSynchronizer * 口絵にカラー版掲載

再生位置の色が変わる歌詞をクリックしてジャンプすることで、楽曲の様々な箇所の歌詞を自在に行き来して能動的に聴くことができる

時に早送り)することができ、自分でサビを探す煩わしい作業から解放される。また、図 1 上側には楽曲中の繰り返し構造を視覚化した「音楽地図」が表示されており、ユーザはこれを見ながら興味のない区間を容易に飛ばすことができる。「音楽地図」の横軸は時間軸で楽曲全体を表示しており、最上段がサビ区間の繰り返し構造を、その下の 5 段が他の様々な長さの区間の繰り返し構造を表している（各段で着色された区間同士が類似していて繰り返しを表す）。従来の音楽再生インタフェースでは、楽曲単位でしか興味のない音楽を飛ばせなかったのに対し、SmartMusicKIOSK によって初めて、楽曲内部の興味のない箇所も容易に飛ばすことが可能になった。

2.2 LyricSynchronizer: 歌詞と楽曲との自動的な時刻同期

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。図 2 のように、既存の楽曲から抽出した歌声とその歌詞のテキストを自動的に同期させて着色することで、楽曲中のどこが再生されているか

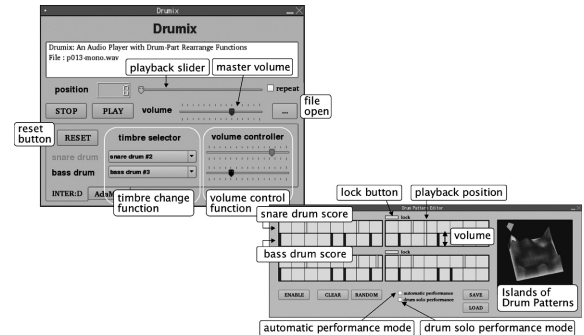


図 3 Drumix

音楽再生中にドラムスの音量や音色、ドラムパターンを能動的に変更できる

が歌詞上で容易にわかるようになる。さらに、画面に表示された歌詞上の任意の単語をクリックすることで、その単語の位置へジャンプして再生することも可能である。

2.3 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

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。図 3 上側のウィンドウにより、楽曲再生中にそのバスドラムとスネアドラムの音量や音色を変更することができる。さらに、下側のウィンドウを使うと、ドラムパターン(ここではバスドラムとスネアドラムの演奏パターン)を容易に編集したり引用したりすることが可能である。このようにユーザは Drumix を用いることで、既存の楽曲の再生中に気ままにドラムスの音色やパターンを変える能動的なインタクションを楽しむことができる。

2.4 Musicream: 楽曲との能動的で多様な出会い方を可能にする総合的な音楽再生インタフェース

エンドユーザが抱く「何か聴きたい」「思い通りに聴きたい」という欲求を満たすことができる音楽検索・再生インタフェースである。これは、従来の音楽再生インタフェースで典型的であった「曲名やアーティスト名を列挙する」という固定概念から脱却し、能動的で自由度の高い音楽再生を実現することを目指したものである。まず、「何か聴きたい」という欲求を満たすために、様々な楽曲に容易に出会うことができる機能を提供する。図 4 右側を楽曲群(円盤)が上から下へ次々と流れ、ユーザが楽曲を何気なく手に取って聞くことで、未知の楽曲に能動的に出会うことができる。その楽曲が気に入ったら、流れている他の楽曲にそれを接触させることにより、雰囲気の良い楽曲だけを選択的に次々とくっつけられる

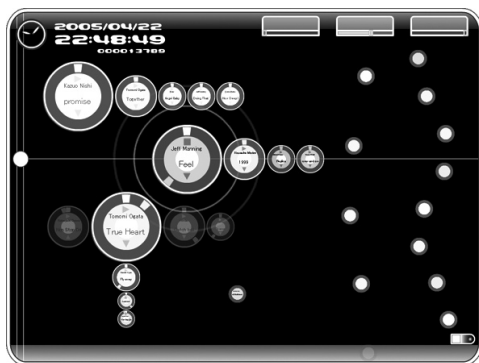


図 4 Musicream * 口絵にカラー版掲載

音楽コレクションをブラウジングして未知の楽曲に能動的に出会うことができる

(制御可能な類似楽曲検索に相当する)。次に、「思い通りに聴きたい」という欲求を満たすために、複数のプレイリスト間の再生順を指定するような自由度の高いプレイリスト編集を可能にする。そして、Musicream 上のすべての操作を記録することで、あたかもタイムマシンを使ったかのように、過去の画面を呼び出したり、そこから新たな未来を作ったりできる。各円盤はその楽曲全体の雰囲気(曲調)を表す楽曲特徴量に基づいて着色し、楽曲間の類似度はその特徴量同士の類似度とした。

3. 音声補完シリーズ

我々は、音声と言語情報と非言語情報を同時に伝達できることが本質的で重要であり、音声の持つ潜在能力を引き出すための鍵が非言語情報にあると考えて、それを有効活用する一連の音声インタフェース機能を「音声補完シリーズ」と称して実現してきた[2]。従来の音声認識研究の多くは、音韻や単語のような言語情報をいかに音声から得るかに焦点を当てて、認識率の向上を競ってきた。そうした技術は重要ではあるものの、認識率を高くするだけでは不十分であり、音声認識を利用したインタフェースはまだまだ使いにくい。そこで我々は、従来は誤認識を招く一因として嫌われていた非言語情報(パラ言語情報も含む)を、ユーザに特定のルールに沿って意図的に発声してもらうことで、従来なかったような使いやすい音声インタフェース機能を実現する新たな研究アプローチに取り組んでいる。具体的な非言語情報としては、書き言葉にはない、音声に特有の言い淀み(有声休止)と声の高さ(基本周波数)の二つを推定する(ここでは有声休止を「母音の引き延ばし」の意味で用いる)。以下では、それらを活用する「音声補完」「音声シフト」「音声スタータ」「音声スポッタ」の四つを紹介する。

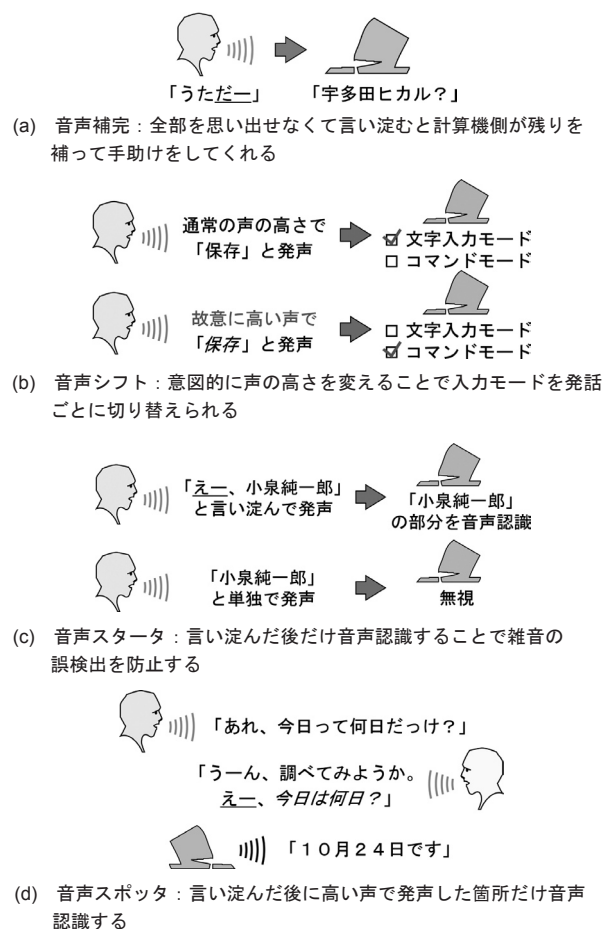


図 5 音声補完シリーズ

3.1 音声補完

音声入力時に、あるフレーズの一部しか思い出せず、に断片だけを言って言い淀むと、音声入力システム側がその残りを補って入力することを可能にする機能である(図 5 (a))。例えば、ユーザが単語発声途中に「うただー」のように有声休止(下線で示す)によって言い淀むと、「宇多田ヒカル」を含む補完候補の一覧(図 6)を見ることができる機能を提供した。これにより、入力中に困って言い淀めば手助けが受けられる使いやすい音声入力の実現できた。

3.2 音声シフト

ユーザが声の高さを意図的に変えることで、音声認識時の入力モードを切り替えることを可能にする機能である。声の高さは通常の音声認識では使われていないが、音声シフトではこれを活用し、普通に発声した発話と故意に高く発声した発話を異なる入力モードに割り当てることで、音声のみでモード指定と情報入力とを同時に行うことを可能にした。例えば、音声ディクテーションに

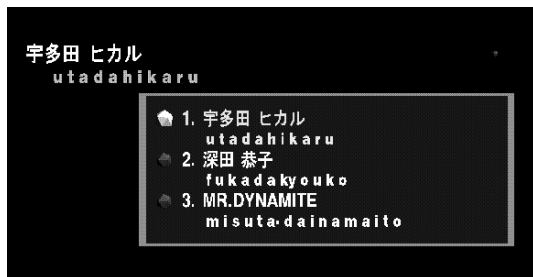


図6 「うただー」と言い淀んで音声補完候補が表示された例

において、「改行」と普通に発声するとその文字が入力され(文字入力モード),「改行」を故意に高く発声(斜体で示す)すると行末が改行される(コマンドモード)機能が実現できた(図5(b)).

3.3 音声スタータ

非定常雑音下で頑健な音声認識を実現するために、言い淀んだ後だけ音声認識することで雑音の誤検出を防止する機能である。従来、雑音下での音声認識では、発話区間の切り出しミスが誤認識に与える影響が大きかった。そこで、雑音下で話し始めるときには、必ず「えー」や「んー」のような有声休止によって発話の頭で言い淀むことをルールとするインタフェースを実現した(図5(c)). 通常、雑音中にはそうした言い淀みに似た音は含まれないため、様々な物音が鳴っても発話区間と誤らずに適切に無視できる。これにより、ユーザは発話開始時点を音声だけで明示的に指示でき、従来、ボタン操作等による指示でハンズフリーの利点が失われていた問題を解決できた。

3.4 音声スポット

通常の会話に現れないような「母音を延ばして言い淀んだ後に故意に高い声で発声された特殊な発話」だけをスポッティングして音声認識対象とみなすことで、マイク入力だけで人間同士の会話中の音声認識対象箇所を同定できる機能である(図5(d)). 例えば、「えー、今日は何日」のように、有声休止の後に入力したい文章を故意に高く発声すれば、計算機が「今日は何日」の部分だけを認識し、その答えを教えてくれる。こうした日時、天気、ニュースの問い合わせに関する情報支援や、楽曲再生の曲名指定等が、会話中に音声だけで実現できた。

4. 議論

こうした一連の音楽・音声に関するインタクションを研究する上で、我々が日頃大切だと思っている考え方を議論する。

4.1 音を効果的に視覚化する

2章、3章で紹介したインタフェースの多くは、音楽・音声を対象としているにも関わらず、視覚的な画面表示を効果的に活用している。これは、目には見えない「音」を扱う上で、その内容に関するインタクションを容易にするためには、「音」だけの提示では困難だからである。また、「音」は静止画と違い、その内容を把握するには、通常はその長さだけ再生時間を要してしまう性質があり、視覚化は大切となる。

特に音楽の場合、楽曲が内在している情報(音楽要素)を適切に視覚化して、インタクションの対象として操作可能にしなければ、何度も再生を繰り返すことになり、効率が悪くなる。再生しながらリアルタイムに操作する場合もあるが、過去に再生した部分を操作しようとする、視覚化なしでは記憶する負担が大きくなってしまう。さらに、より積極的に視覚化を使い、本来は、その音の長さ分の再生時間をかけて再生しないと把握できない内容を、再生する前に理解できるようにすると効果的である。例えば、2.1節では、音楽地図により、楽曲全体の繰り返し構造やサビの位置を見ただけですぐに把握できることで、興味のないところを飛ばして、効率の良い再生を可能にした。2.2節でも、歌詞を手がかりとすることで、全体を再生しなくても、再生位置の変更を可能にした。2.4節では、円盤の色が楽曲全体の雰囲気(曲調)の手がかりとなっており、楽曲を再生する前に、ある程度の選別を可能にした。

4.2 シンプルにした上でユーザに協力を求める

「音楽・音声インタクション」で困るのは、ユーザが、今何をすればよいかわからない状態になってしまうことであり、それを避けるためには、インタクションをシンプルにした上で、ユーザに協力を求めるとよい。そのために、2章、3章のような複数の個別事例として実現しているのは、それらすべてを統合した複雑なインタフェースを目指すよりも、適切だと言える。

特に音声の場合、音声対話は自然ではあっても自由度が高すぎて、インタフェースとして実現されると、ユーザは何を発声してよいかわからなくなることが多い。例えば、3.1節の音声補完では、言い淀めば手助けをしてくれるという、人間同士の対話で経験している自然なルールを取り入れたが、ここで重要なのは、音声対話を直接実現したのではなく、音声対話の一場面をインタフェース機能として適切に抽象化してルール化したことである。結局音声対話であってもユーザに協力してもら

う必要があるのであれば、いっそのこと最初から音声インタフェース機能のルールを明示し、それに沿って発声してもらう方がよい。自然な音声インタフェースということで、音声対話を目指す研究も大切ではあるものの、現状で実用的なインタフェースとして活用してもらうには、他の計算機インタフェースと同様に、固有のルールを明示的にユーザに理解してもらうのが有効である。

さらに、ルールとして明確になっていれば、人間が未経験だった不自然なルールであっても、ユーザは使いやすいく感じることもある。3.4 節の音声スポットは、そうした不自然さを敢えて利用し、人間同士の会話中に通常出現しない発話、つまり、不自然な発話をユーザに言うってもらうことで、人間同士の会話に干渉せずに様々な支援が提供できた。こうした研究は、自然なルールだけに視野を狭めていたのでは、決して生まれなかった。

4.3 「点」だけでなく「面」を作る

ある一つの研究事例で示せる研究思想は限られている。特に論文の場合には紙面が限られていることも影響し、個々の事例でインタフェースとして具体化でき、実証できる範囲は、できるだけ大きくすることは心掛けつつも、限界があるからである。

むしろ、個々の事例を「点」とみなし、複数の事例を積み重ねることで点と点を結んで「線」を作り、さらに様々な方向の線を増やして「面」を作ることで、より大きな思想を伝える研究アプローチが大切だと考えている。点だけでは、その研究の持つメッセージや将来性が伝わりにくい場合もあるが、面にすることで、よりの確に理解され、研究コミュニティで共有されやすくなる。逆に、点の研究から出発するときでも、最初から面としての貢献に発展しやすいよう心掛けている。2 章の「能動的音楽鑑賞インタフェース」や、3 章の「音声補完シリーズ」の一連の事例は、こうした研究ポリシーから生まれてきた。

5. おわりに

本解説では、音響信号の自動理解技術が切り開く新しい「音楽・音声インタラクシオン」の研究事例と考え方を紹介してきた。自動理解技術の進歩は著しいが、人間の理解能力と比べれば、現在の理解技術にはまだまだ制約も多く、性能も不十分である。今後、より多様で高機能なインタフェースを実現するには、自動理解結果に誤りが含まれていても役に立つような応用上の工夫や、音響信号理解技術のさらなる発展が必要である。

2 章で紹介した能動的音楽鑑賞インタフェースの一部は、音楽をより楽しめるようにするだけでなく、音楽をより深く理解できるきっかけを与えてくれる可能性がある。これは「音楽理解力拡張インタフェース」や「音楽理解力向上インタフェース」という新たな研究アプローチに発展していくと考えられる [1]。今後、多くの研究者が様々な形で取り組むことで、日常生活での音楽の聴き方がより多様で豊かに変化していくことを期待したい。また、3 章で紹介した音声補完シリーズは、言語情報と非言語情報の両者を的確に認識して活用していたが、音声はその両者を同時に伝達できることが本質であり、ここにこそ音声インタフェースの優位性がある [2]。今後、音声の持つ大きな潜在能力を引き出す多様な音声インタフェース機能が研究されていくことを期待したい。

謝辞

本解説で紹介した研究事例を共同研究した秋葉 友良氏、伊藤 克亘氏、奥乃 博氏、尾本 幸宏氏、北山 広治氏、後藤 孝行氏、小林 哲則氏、速水 悟氏、藤原 弘将氏、吉井 和佳氏 (五十音順) に感謝する。その研究の一部は、科学技術振興機構 CrestMuse プロジェクトによる支援を受けた。

参考文献

- [1] 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，情報処理学会 音楽情報科学研究会 研究報告，2007-MUS-70-9, pp.59-66 (2007)
- [2] 後藤 真孝：非言語情報を活用した音声インタフェース，情報処理学会 音声言語情報処理研究会 研究報告，2004-SLP-52-7, pp.41-46 (2004)
- [3] 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)

【略歴】

後藤真孝 (GOTO Masataka)

産業技術総合研究所 主任研究員

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)。同年、電子技術総合研究所 (2001 年に産業技術総合研究所に改組) に入所，2005 年より現職。2000 年から 2003 年まで JST さきがけ研究 21 研究員，2005 年から筑波大学連携大学院准教授，2008 年から統計数理研究所客員准教授を兼任。科学技術分野の文部科学大臣表彰 若手科学者賞，ドコモ・モバイル・サイエンス賞 基礎科学部門 優秀賞等 22 件受賞。