

Query by Humming with the VocalSearch System

VocalSearch システムによる鼻歌検索

William Birmingham, Roger Dannenberg, and
Bryan Pardo

作曲者や演奏者、タイトルを知らない？

あなたが知っている主旋律を使って欲しい曲をシステムに探させよう。

アップルコンピュータの iTunes や地元の公共図書館などで楽曲を探す際、通常は、探したい曲についての何らかの文字情報（メタデータ）、例えば、作曲者、演奏者、タイトルなど、を検索エンジンに入力する。しかし、その曲自体は知っているがメタデータは知らない、といった場合には標準的な検索エンジンは選択肢となりえない。その代わりに、曲の一部をハミングしたり口笛を吹いたりすることで、メタデータではなく曲の内容（メロディ）に基づいた検索エンジンにクエリを与えることができよう。歌ったり、鼻歌を口ずさんだり（hamming）、口笛を吹いて与えたメロディに基づいて曲を検索できるシステムは、鼻歌をいつも使うわけではないのだが、鼻歌検索（Query by Humming, QBH）と呼ばれる。

音楽フィンガープリント（Music Fingerprinting）とは、特定の曲を見つけるための関連アプローチであり[2]、そこでは、需要の多い録音の一部を検索例として与える。QBH と違い、音楽フィンガープリントは、元の楽曲の音響的な特性を精密な計測結果を探索鍵、すなわち、フィンガープリントとして使用している。故意に、音楽フィンガープリントシステムは、検索クエリがデータベース中の録音と正確に一致している時にだけ機能する。演奏が異なっている場合（例えば鼻歌で口ずさんだメロディなど）は、フィンガーブ

リントを歪め、マッチングがうまくできなくなる。

音楽専門司書は、QBH の有用性を証言している。というのは、後援者が電話をかけてきて、電話越しに口ずさんだメロディの曲名を見つけてという依頼をよく受けているからである。このような音楽クエリは多くの人々にとって自然であり、検索時点では所望の録音のコピーにアクセスする必要がない。さらに、QBH システムでは、フィンガープリントシステムよりも、より広範囲な結果を得ることができる。というのは、ユーザの心にとまっているのが、メロディであり特定のアーティストでないからである。（例えば、「Jesus Just Left Chicago」という曲の Phish による演奏と ZZ Top による演奏の両方を検索したい」場合など。）

本稿では QBH の説明のために、実際に運用されている例として、ミシガン大学とカーネギーメロン大学が（全米科学財団（National Science Foundation）の資金援助によって）開発した Vocal Search[9] システムを用いる。というのは、本システムが、現行では最高の性能を誇りシステムの1つであり、簡潔な設計とユーザインタフェースを有しているからである（本稿の図参照のこと）。クエリは、通常、ユーザから与えられるメロディ、主旋律（theme）、印象的な旋律（hook）、楽器リフ（riff）、またはその他の曲の覚えやすい部分の形式をとる[9]。したがって、多くの QBH

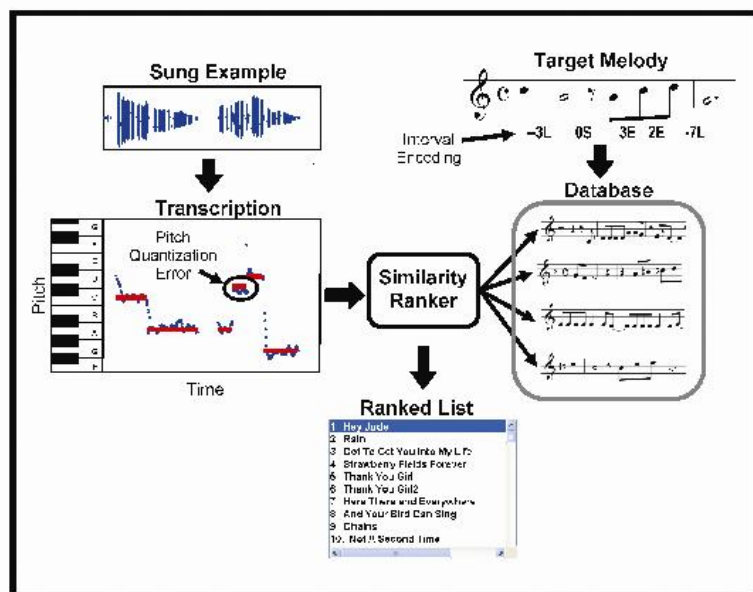
システムは旋律的な音楽を対象として設計されている。リズムや音色が重要視される音楽には QBH 技法は不向きである。もっとも、最近の興味深い複数の研究がリズム認識にも取り組んでいる (www.songtapper.com) [4]。

典型的な QBH メロディ比較技術—文字列対応付け (string alignment), n グラム (n-gram), マルコフモデル (Markov models), 動的時間伸縮 (dynamic time warping)—は、データベース中の単旋律 (monophonic, 単一メロディライン) のメロ

ディを単旋律クエリと比較する。大部分の西洋音楽は調波構造的 (harmonic) か、多重奏 (polyphonic, 同時に複数声部が演奏) であるため、そのような音楽を検索可能にするためには、最初に、音楽から各単旋律メロディを抽出する必要がある。同時に複数の楽器が演奏されるような音響信号に対して標準的な音高追跡手法 (pitch tracking methods) は信頼性が低下するため、これらのメロディは音響信号から直接抽出されるわけではない。この未解決問題は記号符号化された音楽 (例えば, MIDI (musical instrument digital interface) プロトコルファイルなど) を使用することで回避できよう。

MIDI ファイルに含まれているのは、演奏すべき音符、音符の順序、各音符の持続区間と音量についての指示である。MIDI ベースの携帯電話の着メロの普及の (部分的な) おかげで、何万曲もの MIDI ファイルが利用できる。多重奏 MIDI ファイルは、さまざまな方法で、単旋律ファイルの集合に自動的に変換が可能である [8,12]。その結果得られたファイルは、次に、メロディデータベース中の検索鍵として使用できる。

適合率 (precision) と再現率 (recall) を



VocalSearch QBH システムにおけるクエリ処理方法。

向上させるため、QBH システムは、典型的には、楽曲全体ではなく、メロディ主旋律 (例えば、ベートーベン「交響曲第 5 番」の最初の 5 音や、曲の 1 節など) のデータベースを検索対象とする。このようにデータベースを制限—単声の代表的な主旋律—することで、クエリと関係ない楽曲のあいまいな部分が思いがけずマッチ (例えば、クエリ中の下降音階がウォーキングベースライン (walking bass line) とマッチする) 可能性を減らしている。また、検索対象データ量を削減することにより、処理を 100 倍以上も高速化できる。例えば、VocalSearch データベース中のビートルズの曲は、歌詞の主旋律と、コーラス (サビ部分) の異なる主旋律とで表現されている。図に、標準的な音楽表記の形式でこれらの主旋律の例を概説する。

主旋律識別システムの開発は、特別なチャレンジ問題である。主旋律は、曲中のどこでも出現する可能性があり、いかなる楽器や声部で演奏される可能性もある。例えば、ドボルザーク (Dvorak) の「アメリカ四重奏曲

**多くの研究者が、処理時間、再現率、適合率の観点から、
よい性能を示す結果を報告してきた。**

(American Quartet)」では、主旋律がヴィオラで導入されるが、主旋律題材の導入にこの楽器が使われるのは稀である。主旋律は変奏されて繰り返される。その変奏は僅かの時もあれば、そうでない時もある。時には、主旋律が単純で、曲中の非メロディ部分と間違えやすいこともある。歴史的には、主旋律データベースは高度な技能を持つ音楽家が作成してきた。これは、労働集約型作業であり、大規模化するのが難しい。一方、自動主旋律抽出法も有効であることが示されてきた[7]。

そのノイズは何？

クエリの品質も多岐にわたっていた、訓練された音楽家であっても「ましな」クエリをQBHシステムに提示できるとは限らない[9]。人は主旋律を正しく思い出せない場合もあるし、複数の主旋律や声部（リードギターとボーカルなど）をまぜこぜにして混合クエリとする場合もある。また、人はクエリ生成での問題も有している。クエリを提示する際に、望みの音高を歌うことができない場合や、笑ったり咳き込んだりしてしまう場合もある。コンピュータはそのような意図しない人工物（artifact）を「聞いて」しまい、クエリの一部だと判断してしまう。

採譜、つまり音響信号を音楽表記への変換もまた、ノイズの別の源である。採譜者も、人工物を加えたり、音符を落としたり、リズムを間違えたりすることが少なからずある。このように、「完全な」クエリであっても、採譜後には完全ではなくなるように思える。図中の採譜部分は、いくつかのきちんと歌われた音符の採譜がどのように見えるかの例である。図上部の青点は、歌われたクエリから推定された基本周波数の系列を示しており、音高がなめらかに変化しながら揺れている様子に注意。赤線は、採譜者によって推定された音符（音高と持続長）であり、最も近いピアノのキーに量子化されている。この例では、4番目の音符で量子化誤りが発生し、Cである

べき音が **C#** と上がっている。

そのような誤りは、訓練されていない歌手の不正確な音程によるのか不正確な採譜によるのかであるかにかかわらず、採譜されたクエリをデータベース中のターゲットとマッチをとる際に含める必要がある。量子化誤りを扱う1つのアプローチは、音高曲線（pitch contour）を量子化せずデータベース中の主旋律と直接マッチをとることで、誤りを回避することである[5]。VocalSearchは、誤差尤度の事前推定を通じて、採譜者と歌手の誤りを扱う。システムの学習は、既知楽曲のメロディを歌ったデータに対して行う。採譜されたメロディをもとのメロディと比較し、その差異を記録する。学習により、歌手と採譜者の誤りについての確率モデルを構築する。

最良の推量（Best Guess）

多くの研究者がメロディの比較技術に取り組んで [3,6,12]、その一方で、別の研究者が多重奏の比較技術に取り組んできた[8,11]。文献[1]では、QBHのための最もよく使われているメロディの類似度基準が比較分析されている。例えば、n グラム、動的時間伸縮、マルコフモデル、文字列対応付けなどがある。この分野の文献の多くは、www.ismir.net/all-papers.htmlにある音楽情報検索国際会議（International Conference on Music Information Retrieval）の論文レポジトリからも入手可能である。

VocalSearch[9]は、ターゲットとクエリ間の類似度を計算するために、確率的文字列対応付けアルゴリズムを使用している。文字列対応付けは、クエリが各々のターゲットとどの程度類似しているかを、編集距離（edit distance）に基づいて決定する。「コスト」（例えば、変更された文字数）とは、ある文字列（採譜されたクエリ）を別の文字列（データベース中の主旋律）に変換するのに要する手間である。最低コストでクエリに変換できた主旋律が、最良マッチだと判断する。

QBH システムは、何万もの主旋律、つまり音楽検索経験のない人ならほとんどマッチをとるのができないような離れ業、を検討している。

VocalSearch では、これらのコストは前述の学習された誤りモデルに基づいており、そこには、歌い手が歌う際にどのような誤りを犯すかについての情報を含んでいる。例えば、短3度の音程を上げようとして、長3度に上がってしまうことはよくある。しかし、長7度まで上がってしまうことはほとんどあり得ない。このように、今クエリが2つの主旋律、1つが短3度で上がる主旋律、もう1つが長7度で上がる主旋律、とマッチを取ろうとしており、クエリが提示するのが長3度の上がりであれば、前者の主旋律の編集コストの方が低いと判断することになる。

クエリとターゲットに使われるメロディ符号化法は、QBH システムが扱える誤りの種類に大きな影響を与える。例えば、クエリとターゲット間で調の不一致問題をうまく回避するために、ある研究者は絶対音高でなく音程距離 (intervallic distance, 連続する2つの音高間の半音の個数など) を使用した。また、歌声の音高の範囲を考慮に入れるために、通常、主旋律とクエリの表現からはオクターブ情報の多くを破棄し、メロディを1, 2オクターブの範囲に制限する。VocalSearch は、音程距離を使用し、リズムを音符の持続長の比で表現する。図には、ビートルズの "Hey Jude" からとった主旋律の例を示す。隣接する音高間の半音の個数は、整数で符号化し、2つ目の音符の持続長は、1つ目のそれと相対的に、Longer (L), Shorter (S), Equal (E) として符号化している。

まとめ

このような技術的な困難にもかかわらず、QBH システムは目を見張るほどよく機能している。多くの研究者が、処理時間、再現率、適合率の観点から、よい性能を示す結果を報告してきた。例えば、文献[10]では、ビートルズのデータベースを用いて、人間に近い性能が得られたという結果を報告している。QBH システムは、過去数年、音楽ライブラリーにおいてで試用ベースで使用されてきた[6]。本稿の第3著者は、自分個人の音楽コレクションにアクセスするのに、日常的にVocalSearch を使用している。一方、個人の音楽ライブラリー、音楽教育、カラオケシステムのための技術的に興味深い応用が可能に

なりつつある。歌われたクエリを認識する能力はまだ人間が上だが、QBH システムはその性能が向上し続けており、何万もの主旋律、つまり、音楽検索経験のない人であればほとんどマッチをとるのができないような離れ業を検討することが可能である。

文献

1. Dannenberg, R., Birmingham, W., Pardo, B., Hu, N., Meek, C., and Tzanetakis, G. A comparative evaluation of search techniques for query by humming using the MUSART testbed. *Journal of the American Society for Information Science and Technology* 58, 3 (Feb. 2007).
2. Haitisma, J. and Kalker, T. A highly robust audio fingerprinting system. In *Proceedings of the Third International Conference on Music Information Retrieval* (Paris, Oct. 13–7, 2002); www.ismir.net.
3. Hewlett, W. and Selfridge-Field, E., Eds. Melodic similarity: Concepts, procedures, and applications. In *Computing in Musicology*. MIT Press, Cambridge, MA, 1998.
4. Kapur, A., Bening, M., and Tzanetakis, G. Query-by-beat-boxing: Music retrieval for the DJ. In *Proceedings of the Fifth International Conference on Music Information Retrieval* (Barcelona, Spain, Oct. 10–15, 2004); www.ismir.net.
5. Mazzoni, D. and Dannenberg, R. Melody matching directly from audio. In *Proceedings of the Second International Conference on Music Information Retrieval* (Bloomington, IN, Oct. 15–17, 2001); www.ismir.net.
6. McNab, R., Smith, L., Witten, I., Henderson, C., and Cunningham, S. Towards the digital music library: Tune retrieval from acoustic input. In *Proceedings of the First ACM International Conference on Digital Libraries* (Bethesda, MD, Mar. 20–23). ACM Press, New York, 1996, 11–18.

7. Meek, C. and Birmingham, W. Automatic thematic extractor. *Journal of Intelligent Information Systems* 21, 1 (2003), 9–33.
8. Pardo, B. and Sanghi, M. Polyphonic musical sequence alignment for database search. In *Proceedings of the Sixth International Conference on Music Information Retrieval* (London, Sept. 11–15, 2005); www.ismir.net.
9. Pardo, B., Shifrin, J., and Birmingham, W. Name that tune: A pilot study in finding a melody from a sung query. *Journal of the American Society for Information Science and Technology* 55, 4 (Feb. 2004), 283–300.
10. Pardo, B. and Birmingham, W. Query by humming: How good can it get? In *Proceedings of the Workshop on Music Information Retrieval* (Toronto, 2003).
11. Pickens, J., Bello, J., Monti, G., Sandler, M., Crawford, T., Covey, M., and Byrd, D. Polyphonic score retrieval using polyphonic audio queries: A harmonic modeling approach. *Journal of New Music Research* 31, 1 (June 2003), 223–236.
12. Uitdenbogerd, A. and Zobel, J. Melodic matching techniques for large music databases. In *Proceedings of the Seventh ACM International Conference on Multimedia* (Orlando, FL, Oct. 30–Nov. 5). ACM Press, New York, 1999, 57–66.

William (WPBirmingham@gcc.edu) は、ペンシルベニア州グローブ市グローブ市大学 (Grove City College) の計算機科学科の学科長であり、計算機科学と電気工学の教授である。 Roger B. Dannenberg (roger.dannenberg@cs.cmu.edu) は同州ピッツバーグ市カーネギーメロン大学の計算機科学と芸術の准教授である。 Bryan Pardo (pardo@cs.northwestern.edu) は、イリノイ州エヴァンストン市ノースウェスタン大学計算機科学学部助教授であり、音楽学部の併任である。	Birmingham
---	--------------------------

訳：藤原弘将（京都大学・情報学研究科）