

# 音楽情報処理 最前線!

## ■第11回

未来の音楽の楽しみ方、作り方はどう変わるのか？  
コンピュータは音楽を理解できるようになるのか？  
コンピュータを使って音楽を研究する「音楽情報処理」  
という研究分野が、世界的に注目を集めています。  
本連載では、そうした最先端の研究事例を紹介していきます。

### この曲、何だっけ？ 歌で音楽を探す「歌声検索」

曲の一部分を歌い、その音声を使って検索をするのが「歌声検索」。今回は、歌詞とメロディーを同時に使う歌声検索について紹介する。その精度は90%以上だ。

#### この曲、何だっけ？

「昔テレビで聴いた曲、サビしか覚えていないんだけど、誰の何という曲だっけ？」と思った経験がある人は多いのではないだろうか。こういふときに使えるのが音楽情報検索だ。音楽に関する情報をコンピュータに入力すると、関連する音楽の一覧を表示してくれる。

音楽情報検索に使われる情報には、曲のタイトルや歌手などの「メタ情報」と、曲の歌詞やメロディーなどの「コンテンツ情報」がある。歌詞はテキスト情報なので、歌詞データベースがあれば、歌詞検索ができる。メロディーの検索では、たとえばピアノロールでメロディーを指定すれば、それに合う曲を検索するのはそれほど難しいことではない。

#### もっと便利に使いたい

……と、原理的には音楽の検索は難しいのだけれど、音楽をちょっと検索したいときにいちいちピアノロールでメロディーを入力するのは面倒だ。それに、人間が歌を覚えるときにはメロディーに歌詞が乗った状態で覚えていることが多いので、曲の歌詞だけを思い出して入力するのは思いのほか手間がかかる。

さらに、歌詞とメロディーをバラバラに指定したのでは、歌詞とメロディーの結びつきの情報が生かせないので、検索精度が悪くなる。メロディーと歌詞の結びつきを保ったまま入力できる「ボークロイドエディタ」のような入力インターフェースならいいだろうが、これでは入力がかかります煩雑になる。

元が歌なのだから、自分が歌った歌をそのまま使って検索できればいいのに。それを可能にするのが「歌声による音楽検索」すなわち「歌声検索」である。

#### ハミング検索+音声認識 =歌声検索？

自分の歌った声で検索をする技術はすでに開発されている。ハミング検索である。ハミング検索は、「ラララ〜♪」のように歌った歌声からメロディーを抽出し、それを使って音楽を検索する。ピア

ノロールの代わりに、声でメロディーを入力すると考えれば分かりやすい。ハミング検索は、すでに実用化もされている。一方、しゃべった声を聞き取って文字化する技術もすでに開発されている。音声認識技術だ。音声認識の技術も現在かなり進んでいて、朗読音声文字起こしパソコンソフトも売られている。それなら、ハミング検索と音声認識を同時にやれば、歌声検索なんか楽勝じゃないの？

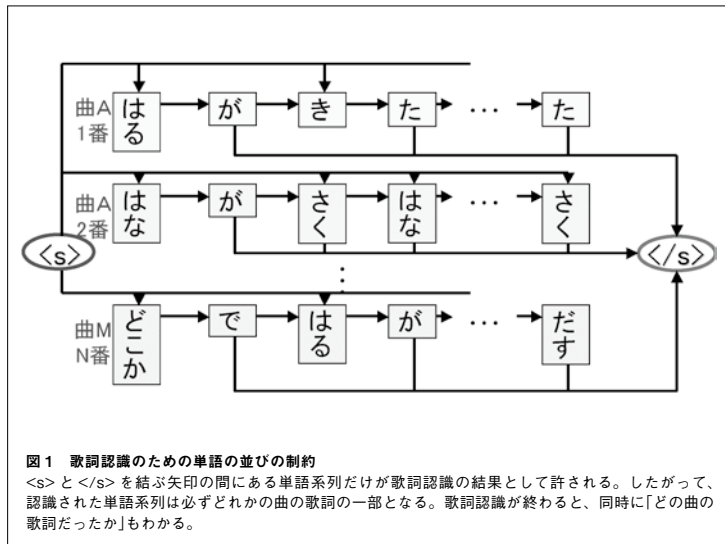
#### 歌声の認識は難しい

ところが、話はそう簡単ではない。歌声の中の歌詞の認識は、通常の音声の認識よりもずっと難しいのだ。そもそも、人間にとってさえ、歌詞の認識は難しい。覚えていた歌詞が実は間違っていた、という経験がある人も多いだろう。これには2つの原因がある。

ひとつは、メロディーに乗った歌詞のリズムやアクセントが、通常の朗読文と大きく違うことだ。

歌の中では、単語の途中の音が長く引き延ばされたり、あるいは単語中の1音内で音程が変化したりする。これらは芸術としての音楽には不可欠なものだけれど、歌を文章として聞き取るためには障害となる。

もうひとつは、通常の文章と比べて、歌詞は単語の間に脈絡がないことが多いことだ。歌詞には拍の数の制限があるので、文章としての自然さよりも拍の数を合わせる方が重視される。そのため、普通の文章よりも、単語の省略や語順の変更が多くなる。音声の認識では語順が重要な情報となるため、歌の聞き取りではどうしても性能が低下するというわけだ。



**伊藤 彰則**  
(いとう あきのり)

1992年東北大学大学院工学研究科博士後期課程修了。工学博士。現在、東北大学大学院工学研究科准教授。人間の声の認識と処理、音声による通信、歌声の認識と検索などの研究をしている。

**鈴木 基之**  
(すずき もとゆき)

1996年東北大学大学院工学研究科博士課程退学。博士(工学)。現在、徳島大学大学院ソシオテクノサイエンス研究部准教授。音声認識、パターン情報処理、データマイニングなどが専門。

**牧野 正三**  
(まきの しょうぞう)

1974年東北大学大学院工学研究科博士後期課程修了。工学博士。現在、東北大学大学院工学研究科教授。音声の認識と処理、音声データベース、画像・映像の情報処理、文字認識など、メディア情報処理を幅広く研究している。

「音楽情報科学研究会」へ参加してみませんか？

情報処理学会 音楽情報科学研究会 (SIGMUS) は、コンピュータと音楽とが関わり合うあらゆる場面を活動対象とする学際的研究会で、年5回の研究発表会を開催しています。研究会に会員登録すると、研究発表会の参加費が無料になるだけでなく、過去の全研究発表会の論文のダウンロードなどの特典があります。研究会の登録方法や研究発表会の開催に関する最新情報などは <http://www.ipsj.or.jp/sigmus/> をご覧ください。

おまえの歌を知っている

我々が開発した歌声検索では、これらの問題を解決するために、次のような仮定を置いた。それは、「ユーザが検索のために歌った歌は、データベースの中にある歌の一部分に違いない」というものである。

一般の音声認識の場合、ユーザは何をしゃべるかわからないので、認識システムはどんな単語の並びでも認識できなくてはならない。これに対して、歌ったものがデータベースの一部であることが仮定できるなら、結果が必ずデータベースの歌詞のどこかになるように認識をすればよいのだ。これは、任意の単語の並びの認識に比べると、かなり強い制約になる。

単語の並びの制約を図1に示す。図の中の四角は単語、矢印は認識結果として許される単語の並びを示す。認識結果としては、<s>の丸から</s>の丸に至る経路の上にある単語の並びしか許されない。図からわかるように、データベースにある歌詞のどこかの単語から始めて、どこかの単語で終わるという認識結果しか出力されないようになっている。このように、認識できる単語の並びを制限することにより、認識精度が向上するだけでなく、認識が終わった時点で「それがどの曲なのか」がわかる仕組みになっている。

でも、これだけでは歌声検索の精度は高くない。なぜなら、似たような歌詞の曲は山ほどあるからだ。例えば「君を愛してる」なんて歌詞が出てくる曲は、世の中には何百曲もあるはずだ。そこでさらに精度を上げるために、認識された歌詞に対応する曲のメロディーと、入力の歌唱のメロディーを比較して、認識結果が本当にその曲なのかどうかを検証する(図2)。もしメロディーが

全然違っていった場合には、認識結果は違う曲である可能性が高いので、その場合には2番目に確からしい認識結果を取ってきて、改めてメロディーを比較し直す。これを繰り返して、「歌詞もメロディーも似ている」候補を探すわけだ。

図3は、この方法を使って検索をしたときの性能を示す。データベースは童謡156曲だ。歌詞認識の結果だけを使ってもそれなりに検索が当たるが、歌詞とメロディーを併用するとさらに性能が上がり、ハミング検索を使った場合よりも高い検索性能となる。ちなみに、歌詞のある歌声は音符に区切るのが難しいので、メロディーだけを使うと全然性能が上がらない。

いつになったら使えるの？

このように高い性能で歌声検索ができるようになったのだが、課題もある。最大の課題は、検索が遅いことだ。

これまでの実験ではデータベースがわずか150曲程度だったが、実用になるデータベースでは、曲数が二桁以上多い。そうなったとき、現在の方式では遅すぎて実用にならないので、検索を高速化する工夫が必要だ。これについては、今後もさまざまな工夫を行っていきたい。

参考文献

M. Suzuki, T. Hosoya, A. Ito and S. Makino, "Music Information Retrieval from a Singing Voice Using Lyrics and Melody Information," EURASIP Journal on Advances in Signal Processing (2007) doi:10.1155/2007/38727

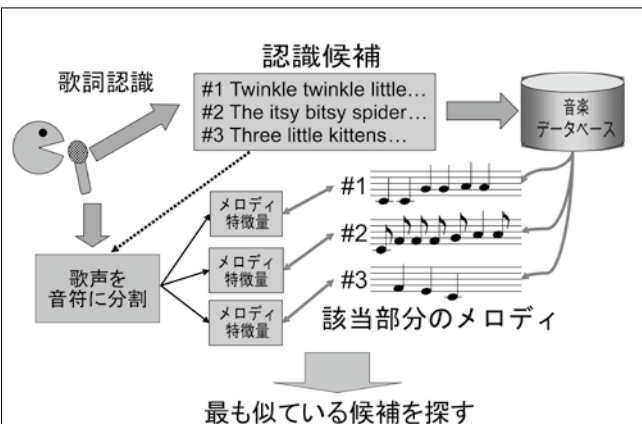


図2 対応メロディーによる認識結果の検証  
歌詞を認識した後、その歌詞に対応するメロディーをデータベースから取り出してきて、それを歌声のメロディーと比較する。歌詞とメロディーのどちらも似ている候補が最終的な検索結果となる。

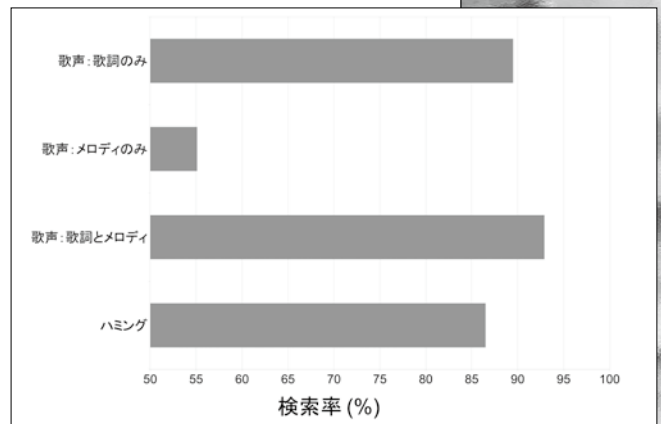


図3 検索結果  
歌詞つきの歌声から歌詞、メロディーだけを使って音楽検索をした場合と、歌詞・メロディーの両方を使った場合の性能比較。グラフの数字は、「最も似ている曲」としてシステムが提示した曲が当たっている割合だ。歌詞とメロディーの両方を考慮することで性能が上がる。歌詞を使わない「ハミング検索」よりも高性能。