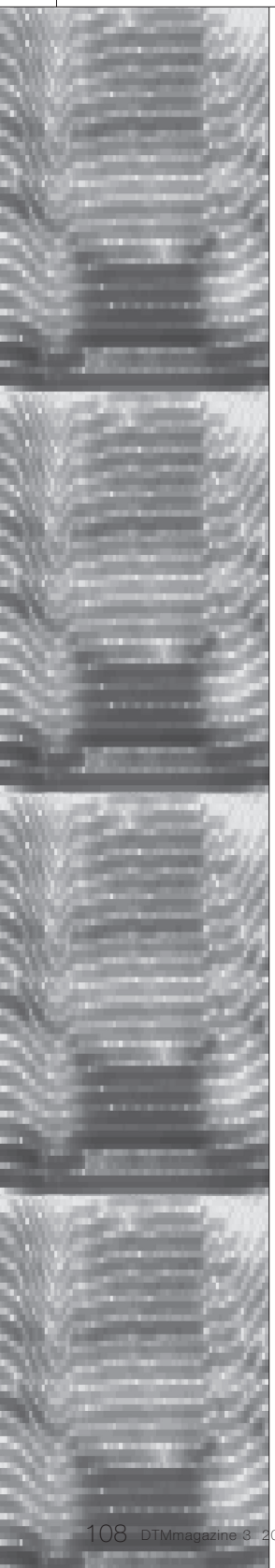
音楽情報処理最前線！

**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。**

「歌声」と「話声」はどう違うか？　人間の“声”を理解するコンピュータの実現を目指して



**人の声には色々な形がある。言葉を伝えたり、笑ったり、泣いたり、歌ったり。
そんな声の違いを聞き分けるコンピュータがあれば、人と同じようにコミュニケーションできるかもしれない。
ここでは、その中でも歌声と話声に注目して、その違いを聞き分けるシステムを紹介する。**

1.「歌声」とは何か

歌声と話声は人間の声という意味では全く同じである。しかし、声を少し聞けば、相手が歌っているのか話しているのかはすぐに分かる。では、人間は声の中のどういった特徴を手がかりに、歌声と話声を聞き分けているのだろうか？

このような疑問から、まず人間の歌声と話声の聞き分け能力を調べてみた。

最初に音声ファイルを用意した。これは、さまざまなジャンルの曲(日本語および英語)を、さまざまな歌唱力の男女100人に歌ってもらい、またその歌詞を読み上げてもらったものだ。

この音声ファイルを短く切り出して10人の被験者に聴いてもらい、その声が「歌声」なのか「話声」なのか二択で選んでもらった。

結果が 図1である。図を見れば分かるとおり、人間は0.5秒でも90%、1秒程度聴けば歌声なのか話声なのか、100%識別できるようだ。

1秒程度で被験者10人全員が正しく聞き分けられるということは、声の中に含まれる「歌声らしさ」もしくは「話声らしさ」は、音楽のジャンルや歌唱力、歌詞の言語などにはほとんど依存していないということになる。

そこで今度は1秒の声の中の、どういった音響的な特徴を聞き分けの手がかりとしているのか調べてみることにした。

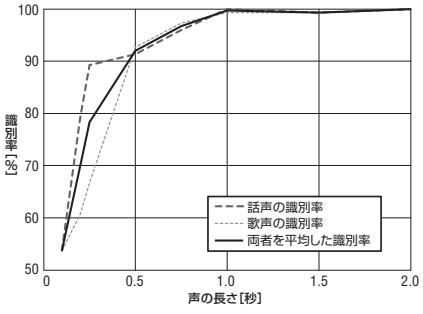


図1　さまざまな長さの歌声と話声を人間が聴いて判断する場合の識別率：歌声の識別率は歌声を聴いたとき歌声と正答した割合、話声の識別率は話声を聴いたとき話声と正答した割合、さらに両者を平均した識別率を示す。

2.「話声」との違いは？

「声」に含まれる特徴として、最も重要なのは声の高さと音色だ。

「歌声」について考えてみると、声の高さの変化は「メロディ」を表すものであるから、聞き分けには特に重要だと考えられる。

また「歌声」の音色には「歌手のホルマント」と呼ばれる特徴的な周波数の「響き成分」(本連載の2008年11月号で紹介)が含まれていることも分かっている。

では、この「メロディ」や「響き成分」を無くしたら、人間は正しく歌声と答えられるだろうか？

これを調べるために、実際に韻律(声の大きさや声の高さの変化やリズム)を崩した声を作成してみた。具体的には、1秒の声を8等分し、これらをランダムな順番でつなぎ合わせたものだ。

一方で、音色を崩した声も作成した。こちらは中高域の周波数成分をローパスフィルタで取り除いた音声ファイルだ(ただし、フィルタのカットオフ周波数を800Hzに設定して、声の高さの変化は聞こえる)。

これらの声を、先ほどと同じ10人に聴いてもらい、その声が「歌声」なのか「話声」なのか二択で選んでもらった。

図2を見ていただきたい。歌声の識別率が大きく下がっているのが分かるだろう。やはりメロディや響き成分を崩した歌声を聴いたとき、人間は正しく歌声と答えられない。

図を使って、その原因を考えてみよう。図3は、韻律を崩した歌声と話声から、声の高さを取り出したものだ。

歌声の特徴であるメロディ(階段状に変化する音程)が崩れてしまい、話声の声の高さの変化と、ほとんど見分けがつかなくなるのが分かるだろう。

図4は音色。破線で囲った響き成分が取り除かれると、歌声のスペクトログラムは話声のスペクトログラムとほとんど見分けがつかないのが分かる。

予想どおり人間は歌声と話声を聞き分けるために、韻律と音色の特徴をどちらも重要な手がかりとしており、また図2の識別率の下がり具合から考えて、韻律がより重要だと考えられる。

一連の聞き取りテストをhttp://www.sp.m.is.nagoya-u.ac.jp/~ohishi/listening_test.html上で体験できるので、是非とも聴いて確認してみて頂きたい。

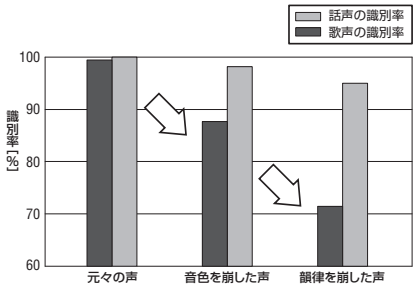


図2　韻律や音色の特徴を崩した声を人間が聴いて判断する場合の識別率：特徴を崩した歌声の識別率が話声に比べて大きく下がった。

大石康智
（おおいしやすのり）

2006年名古屋大学大学院情報科学研究科修士課程修了。現在、同大学院情報科学研究科博士課程に在籍。歌声の音程、音色、音量に含まれる歌い手の個性を分析し、それをモデリングする研究に取り組んでいる。個性を活かした歌声の合成や個性に基づく楽曲検索といった新しい歌声システムの開発を目指している。

後藤真孝
（ごとうまさたか）

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）。現在、産業技術総合研究所主任研究員。計算機によって実世界の音楽・音声コンテンツを自在に扱える技術の確立を目指し、音楽・音声の音響信号の自動理解と、それに基づくユーザインタフェースの研究を中心に、様々な研究課題に取り組んでいる。

武田一哉
（たけだかずや）

1985年名古屋大学大学院工学研究科修士課程修了。博士（工学）。現在、名古屋大学大学院情報科学研究科教授。音声による対話、歌唱、自動車運転など、人間行動の原理と多様性を、信号処理手法により解明する研究に取り組んでいる他、利用者が好きな聴取位置を選択できる、自由聴点オーディオ技術の開発も進めている。

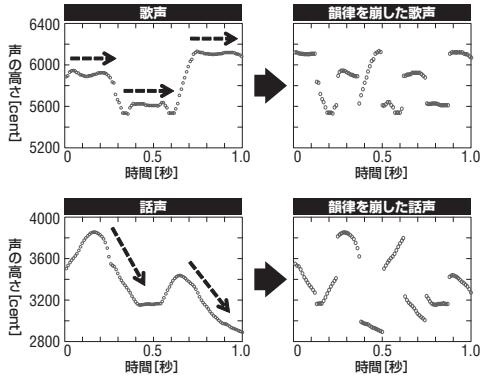


図3　韻律を崩した歌声と話声の声の高さ：階段状に変化する歌声のメロディと話声の右下がりな韻律が崩れてしまい、どちらも声の高さが飛び飛びに変化する。

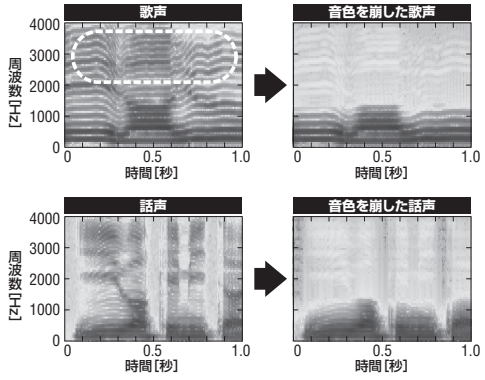


図4　音色を崩した歌声と話声のスペクトログラム：歌声の「響き成分」が取り除かれて、話声のスペクトログラムと見分けがつかなくなってしまう。

3. コンピュータに聞き分けさせる

今回分かったことを活かして、歌声と話声を聞き分けるシステムを作り上げたので、以下にその作成手順を説明しよう。

まず声を周波数分析して、声の高さ（F0情報*）と音色（スペクトル情報）を取り出す。

次に、これらがある時間にどれだけ変化したかを分析して、韻律を表す「F0変動情報」と「スペクトル変動情報」を取り出す。

例えば図3に示す矢印の傾きは「F0変動情報」を表している（ここでは声の高さではなく、声の高さの変化を取り出す必要があることに着目してほしい）。

あとは、このような声の情報を、人間の聞き分け能力を調べたときに使った大量の歌声と話声から取り出して、歌声と話声でどのように声の情報が異なるのか、「確率」を使って事前に学習しておく。例えば、図3に示すように、歌声は音を一定に伸ばすことが多いため、F0変動情報（矢印の傾き）が「0に近い値」となることが多い。

一方、話声は声の高さが文末に向かって徐々に下降するため、F0変動情報は「負の値」となることが多い。

したがって、F0変動情報が0に近い値であれば「歌声である確率」が高く、負の値であれば「話声である確率」が高いということを学習する。新たに未知の声がシステムに入力されたとき、事前に学習したこのような確率に基づいて、歌声なのか話声なのか識別結果を決定するのである。

システムの詳細については、文献[1]を参照していただきたい。

4. あらゆる声を聞き分けさせたい

こうして作ったシステムでの識別結果を図5に示す。今回のシステムでは、人間が100%識別できた1秒の音声で識別率＝約80%、声の長さが長くなるにつれて、識別率はだんだん上がっていくものの、人間の聞き分け能力には追い付けなかった。

しかし、声の中には、まだこのシステムで扱っていない情報が含まれており、これらを活用することで認識率をさらにアップすることが可能だと考えられる。

例えば歌声のリズム感もその一つだ。今回の実験では、ラップやヒップホップの曲を歌った歌声が「話声」と識別されてしまうことがあったが、これらは「リズム」や「声の大きさ」の情報を識別の手がかりに利用することで解決できるだろう。

また「えーと…」とか「あー」のような、言い淀む話声が「歌声」と誤って識別されてしまっていた。このような歌声と話声の中間的な声を分析することも重要な課題である。

さて、今回は歌声と話声に注目したが、人間の口から発声される声には、他にも笑い声や泣き声、囁き声やため息など色々な形態がある。今後はこのようなさまざまな声の聞き分けに関しても研究してみたい。

寂しそうな声だったら、元気づける会話をしてくれたり、楽しそうな声であったら、一緒に笑いあえるような、人の気持ちのわかるコンピュータを実現させたいと考えている。

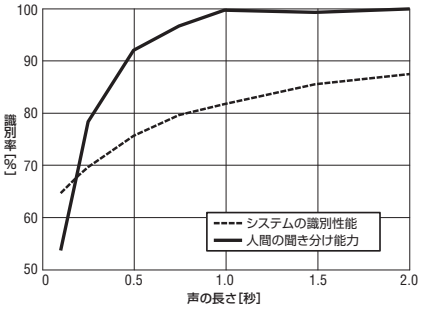


図5　さまざまな長さで切り出した声に対するシステムの自動識別結果：システムは2秒の歌声と話声を87.3%識別できた。しかし、人間の聞き分け能力と比較するとシステムの識別性能はまだまだ低い。

※音響信号処理分野では、声の高さを基本周波数と呼び、更にこれをF0と表記することが多い。

参考文献

[1] 大石康智，後藤真孝，伊藤克亘，武田一哉．スペクトル包絡と基本周波数の時間変化を利用した歌声と朗読音声の識別．情報処理学会論文誌，vol.47, no.6, pp.1822-1830, 2006.