

音情報の視覚化による音ファイル判別ツールの開発

Development of Software for Discriminating Sound File
by Visualization of Sound Information

藤原 成史† Masafumi Fujiwara
 武藤 剛‡ Takeshi Muto
 小宮山 摂‡ Setsu Komiyama

1. はじめに

パソコン、インターネットの普及率の上昇に伴い、インターネット上で音声を含む動画コンテンツの配信・共有が普及し、人々にとってパソコンでの音管理・利用が一般的なものとなってきている。音、動画といった情報がパソコン内に蓄積される様になり、CD、ビデオテープといった記憶媒体を利用する事が少なくなった。パソコンでの管理により、記憶容量が増大し、また、音をファイルとして管理する事によって、CD、ビデオテープと比べて音データの管理が容易になった。しかし、データの中でも音響データは、音を聴くという事をしない限り、名前、サイズ、種類でファイルを判別しなければならない。また、音を聴いて認識する手段は、認識するのに音ファイルの長さ分だけ時間が掛かってしまう。以上のように、PCによって膨大な音データを管理することが出来るようになったが、データが増えていくにつれて、自分の求めている音を判別するのは時間が掛かり困難なものとなっている。

データのファイル数が多い場合、開かなくてもファイルの中身を一覧出来るサムネイルを用いた情報視覚化手法が有効である。

画像ファイルは図1のように、サムネイル部分にファイル内の画像を載せる事で、ファイルを開かなくても内容を理解する事が出来る。また、サムネイルは一度に複数の画像を閲覧する事が出来るので、画像ファイルに対して、サムネイル表示がファイルの一覧性・弁別性の向上に繋がっていると考えられる。



図1 画像ファイルのサムネイル表示

しかし、音ファイルは聴覚から情報を得るものなので、視覚的にファイルの一覧性・弁別性を達成させる事は難しいと考えられる。音ファイルの内、音楽ファイルについては画像ファイルのようにアルバムのジャケットを視覚情報として提示する事でファイルの一覧性・弁別性の向上に繋

† 青山学院大学 理工学研究科 理工学専攻

‡ 青山学院大学 理工学部

がっていると考えられている。しかし、アルバムジャケットは市販の音楽ファイルだけに付加された視覚情報であって、音ファイル全般に共通する視覚情報ではない。

このように、音ファイルに対して、画像ファイルのように視覚情報を提示させ、一覧性・弁別性を促す方法が現状では確立されていない。

2. 音の視覚化方法

情報視覚化は、大量の抽象的な情報を効果的にPC画面に表示することにより、ユーザが情報を理解したり操作したりすることを助けるという概念のことである。従来、シミュレーションによる科学技術計算の結果を画面に表示するサイエンティフィックビジュアライゼーションのことを指すのが一般的であったが、個人が大量のデータを扱う機会が増え、情報視覚化技術に対する需要が高まっている。情報視覚化システムの場合は、一般ユーザの使用が前提であり、またユーザが対話的に表示を変化させることにより静的な表示では得られない情報も得られるようにするために、視点や検索条件などをユーザが変化させることにより対話的に表示を変化させることが出来るようになってきているものが多い。対話的な情報視覚化システムでは、表示すべきデータがユーザ操作により動的に変化するため、表示すべき情報を画面上のどの位置にどのような形・色で表示するかを高速かつ効果的に計算する手法が必要である。現在、一般的なPCでは、図2のようにファイルのアイコン化という、単純な視覚化手法を用いたGUIツールが使われている。総合的にみて現状のシステムよりはるかに優れ、かつ既存のシステムと互換性がなければ新しい視覚化手法を一般に普及させることは難しいと考えられる[1]。



図2 ファイルのアイコン化

従来の音の視覚化方法として、音の大きさの変化をみる事が出来る時間波形と、周波数解析による視覚化がある。図3は女性の声(「また明日」と言っている)を、ハニング窓を用いて時間周波数解析した結果を表している。パワースペクトラムを色によって表現しており、青色から赤色にかけて、パワースペクトラムの強さを表している。この視覚情報により、音の出現位置、周波数分布、周波数の強さ(パワ

ースペクトラム)を、音を聴かなくても理解することが出来る。しかし、図4で見られるように、周波数解析結果をサムネイルとして表示させると、画像の領域が限られているため、視覚情報を正確に認識する事が出来なくなる可能性が出てくる。そこで、情報を選択的に表示する技法が必要と考えられる。

本稿では、音ファイルに関するファイル特定の情報不足を補うため、一般的に音を視覚的に捉える手法である音の時間波形を表示する。また、音波形だけでは情報として少なすぎると考えられるため、人の声の有無は重要な情報であると考え、音分野の中で比較的技術が進んでいる音声・非音声判別技術を用いて音声・非音声区間を検出し、時間波形に情報を付加させて表示する方法を試みる。情報付加の方法としては、音波形に色をつける方法(以下、色つき波形)を取る。2種の視覚情報が聴覚による印象とどの様に対応しているのかを評価する事で、本手法による音情報の視覚化が音ファイルの一覧性・弁別性に有効かどうか検証する。そして、その結果を基に、音ファイルのサムネイル部分に視覚情報を自動的に与え、ファイルの一覧性・弁別性の向上を促すツールの開発をおこなった。

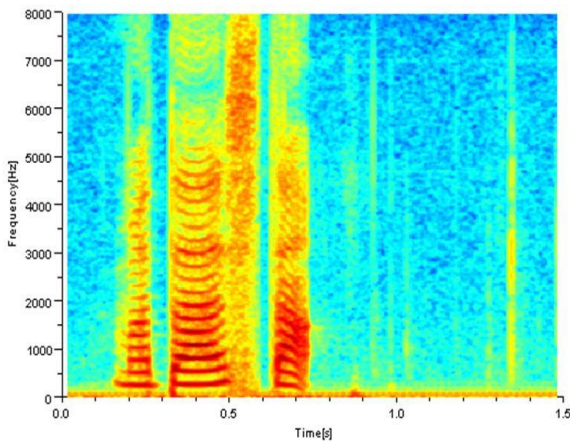


図3 時間周波数解析結果
(横軸：時間, 縦軸：周波数, 色：パワースペクトラム)

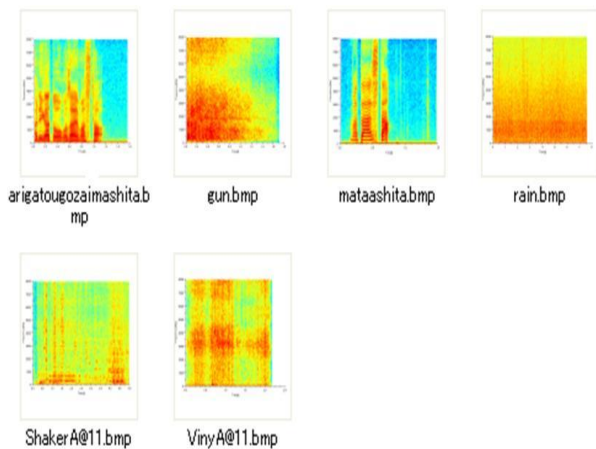


図4 周波数解析結果をサムネイル表示した図

3. 音声・非音声区間検出と表示方法

今回、音声・非音声区間の検出には図5で示すように、音声・非音声のガウス混合分布モデル(GMM)[2]を作成し、それらのモデルと入力音声フレームとの実際の観察データが合っているかどうかを考えるため、尤度計算からの判定で音声・非音声区間を求める[3]。具体的には、入力音声フレームからメル周波数ケプストラム(MFCC)という特徴量を抽出し、その特徴量と、あらかじめ作成しておいた音声・非音声モデルに対して、尤もらしさを表す尤度を計算し、累積させたフレームの尤度比(=音声尤度/非音声尤度)から閾値処理をおこなうことで音声・非音声区間を検出する。そして、得られた音声・非音声区間を時間波形に色をつけることで視覚的に表示した。

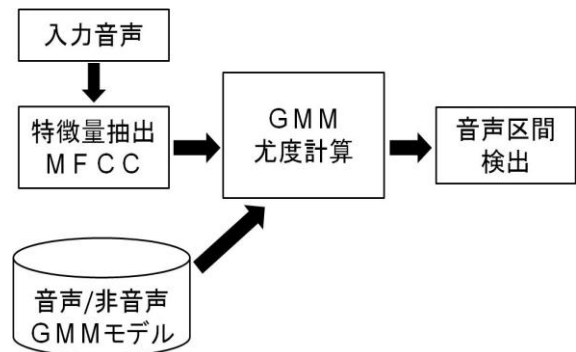


図5 音声・非音声区間検出方法

4. 聴感と視覚情報の対応についての評価

実際にサムネイルを表示するツールを作成するにあたって、今回の視覚表示と聴感との対応関係の評価するための実験をおこなった[4]。実験手順は音を聴いて、音情報をサムネイルに表示させた画像との対応度を5段階評価(不適1~5適)で評価させる。この評価を単色の音波形を表示した場合と、音声・非音声区間を時間波形に色分けした表示の場合でおこなう。被験者は20代の大学生(男女)10人で、用意した音は音声・非音声のみ、音声と非音声が混ざり合った音、音声の後に非音声、非音声の後に音声が出る音、計15種類の音を用意した。この実験から、聴感と時間波形のみでは情報が少なく、十分に聴覚情報との対応がとれないが、音声・非音声区間を時間波形に色分けした視覚情報を与えることで、聴覚情報との対応が大きく改善されることが分かった。因子分析により、聴覚情報と視覚情報の対応は音声の有無に加えて、音声が出現する時間構造が主要な手掛かりとなっていることが示唆された。

5. サムネイル表示ツールの開発

5.1 装置概要

今回提案するシステムは図6,7のように構成されている。まず、ユーザが任意のフォルダを指定し、指定したフォルダに存在する音ファイルに対して、波形のみと色つき波形の2種類のどちらの情報を視覚表示したいのかを選択してもらう。選択された要素に合わせて描画処理をおこない、サムネイル形式によって視覚情報を表示する流れとなっている。

対応する音ファイルの形式は、サンプリング周波数 16[kHz]、モノラルの wav ファイルを対象としている。音声・非音声判別に関して、GMM に必要な音声・非音声モデルの学習データとして、前述の実験環境と同じ設定にするため、音声 GMM には 20 代の大学生(男性)1 人の合計 2 分間の発話データを、非音声 GMM にはホワイトノイズを使用した。先行研究[5]に基づき、音ファイルの 1 フレーム 50.0[ms]ごとに尤度を計算し、4 フレーム間の累積尤度から尤度比を求め、閾値(0.25)処理によって音声・非音声判別をおこなった。閾値よりも尤度比が小さい場合、累積フレームの区間は音声区間となり、閾値の方が大きい場合は非音声区間となる。これにより、音ファイルに対して 200[ms]ごとに音声・非音声判別を行うことができる。

波形の描画として、コマンドライン上で 2 次元、または 3 次元のグラフを作成するためのツールである「gnuplot」を用いた。表示する波形画像の時間軸を揃えるため、最大 5 秒までの波形画像を作成するようにした。5 秒以上の音ファイルに対しては、最初の 5 秒までの時間波形を作成する様にした。

音声・非音声判別に関する処理として、音声認識システムの開発・研究のためのオープンソースの高性能な汎用大語彙連続音声認識エンジンである「Julius」の機能にある GMM 処理を用いた。

サムネイル表示以外の機能としては、表示されたサムネイルを選択すると、そのサムネイルに対応する音ファイルを再生させる機能をつけた。

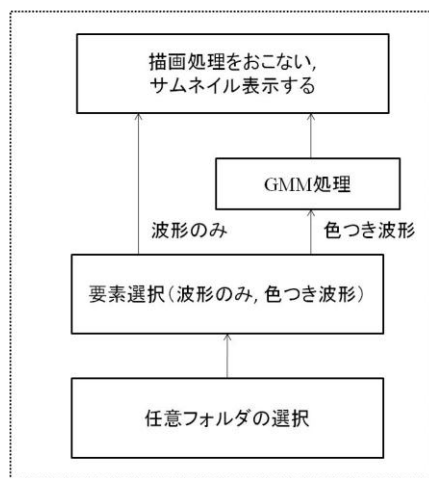


図6 サムネイル表示ツールの流れ

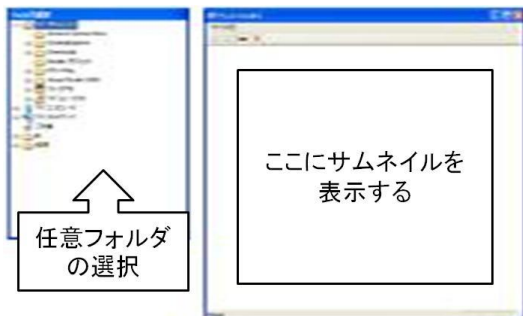


図7 実装の様子

5.2 ツールについての評価

今回作成したツールを実装した際に作成される、波形のみ、色つき波形に関する波形画像(椅子に立ち上がった後、男性の「行ってきます」という声が入った音ファイル)を図7に示す。そして、フォルダ内の音ファイルをサムネイル表示させた結果を図8,9に示す。色つき波形の表示について、赤色が音声区間、青色为非音声区間を示している。

処理に掛かる時間に関して、波形のみの表示では、20個の音ファイルに対して10秒程度で処理をおこなうが、音声・非音声判別処理をする色つき波形では、処理終了までに2分程度時間が掛かった。音声・非音声判別処理の方法で、サムネイル表示に時間が掛かる原因としては、GMM処理による尤度計算が考えられる。処理時間の短縮のために、尤度計算のフレーム幅を広くするといったGMM処理の簡略化を検討する必要がある。

音声・非音声判別処理の精度に関して、音声と非音声混ざった音区間を、必ず音声か非音声かの2種類で分類しなければいけないという事から情報の欠落が生じていると考えられる。尤度計算データは連続的に取れているので、情報の提示方法について検討する必要がある。

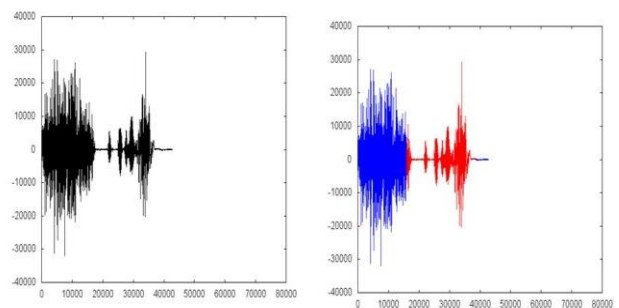


図8 波形画像

(左: 波形のみ, 右: 音声・非音声判別による色分け波形)

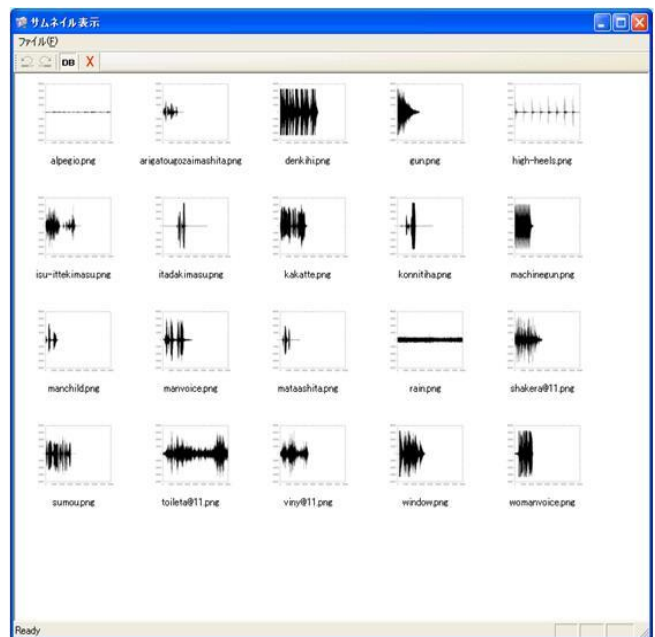


図9 波形のみのサムネイル表示

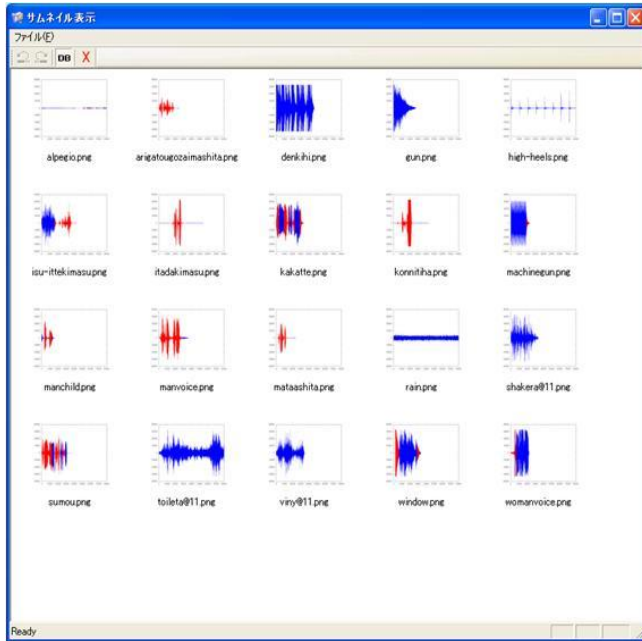


図10 色つき波形のサムネイル表示

6. おわりに

本研究では、音ファイルの一覧性・弁別性の向上のために、音波形に、音声・非音声区間情報を色付けによって重畳したサムネイル表示手法の有効性を明らかにした。それをもとに自動的にサムネイル表示を行う事が出来るアプリケーションを構築した。今後は、今回作成したツールを基に、表示するサムネイルをどこまで小さくできるかといった視覚情報の提示方法について考えていきたい。また、今回は、音声・非音声を2値化して表示したが、音声・非音声の分類に連続性を持たせ、視覚化も連続的な色変化とすることも考えられる。さらに、音声・非音声判別だけではなく、それ以外の情報、たとえば音楽／非音楽判別を提示した場合、音ファイルの一覧性・弁別性の向上にどの様に寄与するか検討する必要がある。

参考文献

- [1] 増井俊之, 情報視覚化の研究動向, ソニーコンピュータサイエンス研究所
<http://pitecan.com/articles/HIS/InfoVis/>
- [2] 山本幸一, Jabloun Firas, Reinhard Klaus, 河村聡典: 識別的特徴抽出に基づく音声区間検出の検討, 情報処理学会研究報告, SLP, 音声言語情報処理, pp.181-186 (2005) .
- [3] C.M.ビショップ: パターン認識と機械学習 上 ベイズ理論による統計的予測, シュプリンガー・ジャパン (2007) .
- [4] 藤原成史, 武藤剛, 小宮山撰: 音声・非音声判別による音ファイルの視覚化, 音講論集, 日本音響学会 (2010, 9) .
- [5] Takayuki Arakawa, Masanori Tsujikawa, 音声検出システム, 音声検出および音声検出プログラム