1P-07

鼻歌検索のための音声特徴量抽出に関する考察

張　夢堯，成　凱
（九州産業大学）

1　はじめに

現在、大量の楽曲がインターネットで利用される．その検索手段はアーティスト名や曲名などの単語によるものが主であるが、曲名などのメタ情報検索時に必ず思い出すとは限らない．メタ情報以外の検索手段として鼻歌検索は有力である．鼻歌検索においてはボーカルの歌声によるメロディが検索対象となる場合が最も多い．ボーカルの歌声を直接にメロディ検索データとして使うことが少なく、ボーカルの歌声には他の楽器にはない性質を基づいて、音声特徴量を抽出してからメロディ検索データとして使うことが多い．

ボーカルの歌声の性質といえば、楽器音は一定の音高のピッチを保つ傾向にあるのに対し、ボーカル音声は常にピッチが変化する傾向にある．ボーカルの音声はピッチが急激に変化することは少なく、なめらかに変化するなど、いろいろな他の楽器にはない性質が持っている．これらの性質をメロディ推定に用いることができると考えられる[1]．

本研究では、ボーカルの歌声の性質を利用して、鼻歌検索の参照メロディデータとして用いることを主目的にしたボーカル特徴量を抽出手法について考察した．

2　先行研究

ボーカル特徴量を抽出手法に関する初期の研究は楽譜情報やMIDI情報を検索対象としていた．しかし、レコーディングされた楽曲に比べて利用できる楽譜は少なく、検索対象としては音響信号のほうが需要が高い．音響信号を対象として扱うために、中間的な情報を用いるやり方がある．メロディらしさのパターンを抽出し、抽出した全てのパターンを用いて入力クエリとマッチングする方法、解析セグメントごとに複数高音の候補のセットを抽出する方法を提案されている．一方、音響信号から直接メロディ情報を抽出してMIDI情報と同様に扱うことによって鼻歌検索を行う方法も考えられる．メロディの推定自体が難しい問題であり、多くの研究者がこの研究に取り組んでいる．ほとんどのメロディ推定手法は複数のピッチ候補を抽出しそれらの中から最適なメロディラインを選択する．研究者が複数のトラッカを用いて有力な候補から主メロディを抽出する方法を提案した．また、必ずしもメロディが他の楽器音と比べて目たつとは限らないため、研究者がボーカル音声をパワースペクトルの時間方向と周波数方向の変化量の違いを用いて抽出したり、ピッチの大まかな変化をピッチトレンドとして用いることで[1]メロディ推定精度を向上させたりした．

3　音声特徴量

音声特徴量は録音した音声に対して前後の無音区間を切り、音声区間のみに対して、音声認識で一般的に用いられる分析手法である．人間が音を聞く仕組み (方向、高さ、大きさ、音色など)の解明、データの符号化、圧縮への応用などの統計的分析、音声認識、音声インターフェースの作成、楽曲のジャンル推定、楽曲検索、推薦などの機械学習、パターン認識における広く応用されている．しかし、収録条件、個人差、データ差などの違いが大きくて本質的な部分がわかりにくいし、データ量や計算量が多くて取り扱いにくいため、生のデータそのものを利用するのは無理がある[2]．ですから、より無駄が少なく、データの本質をあらわした良い特徴量を抽出する必要がある．

音声特徴量は、RMS、MFCC、Spectral Similarity、log-mel spectrum、残響時間、基本周波数、歪み率など[2]、いろいろな種類があって、それぞれの適切な方面で応用されている．そのうちに鼻歌検索のために、音色、音色の時間変化、 高さ、テンポ、リズム、高さなどの音の性質に基づいて特徴量を抽出するMFCC、log-mel spectrum、基本周波数は音声特徴量として最も使われている．

3.1　MFCC

MFCCは聴覚フィルタに基づく音響分析手法で、主に音声認識の分野で使われることが多い．人間の低音に敏感で高音に鈍いという特徴を考慮しつつ、声のスペクトル包絡 (スペクトルのなだらかで少なめの変化)、スペクトルの微細構造 (スペクトルの細かい微妙な変化)を分離したものである．要は、長期的な変化と短期的な変化を分離することで人間の声の変化を見ようとする算段である[3]．

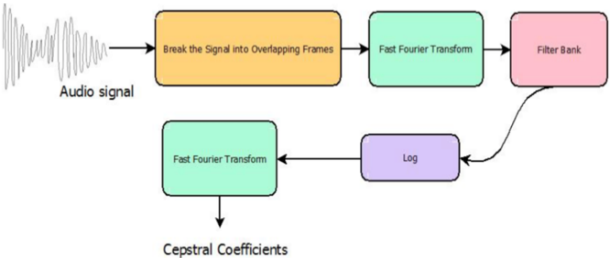

図 1: MFCC の導出手順

MFCCの導出過程は次のようにして行う．

1. 前処理
   (1) プリエンファシスフィルタで高音域強調
   (2) 窓関数をかける
2. 高速フーリエ変換により振幅スペクトルを求める
3. メルフィルターバンクを適応させる

4. 離散コサイン変換
5. 低次元成分を抽出して声道のスペクトルとして音声認識に利用

音声特徴量抽出の手法として、MFCC のメリットは低次成分で個人差の大きいピッチ成分を除去して、音韻の特定にとって重要な声道の音響特性のみを抽出できる．高次成分で声道の音響特性を除去して、ピッチ成分を抽出できる．しかし、デメリットとして雑音のスペクトルが特定の帯域に集中している場合、ケプストラムのすべての係数に影響を及ぼす．

3.2 log-mel spectrum

log-mel spectrum は MFCC の離散コサイン変換無いバージョンである[3].

log-mel spectrum 抽出手順は次のようにして行う．

1. 前処理
   (1)プリエンファシスフィルタで高音域強調
   (2) 窓関数をかける
2. 高速フーリエ変換により振幅スペクトルを求める
3. メルフィルターバンクを適応させる

離散コサイン変換の過程がないため，log-mel spectrum は MFCC に比べて計算コストが低く、計算速度が上がっている．現在では log-mel spectrum を用いることが主流になっている．

3.3 基本周波数

基本周波数とは、信号を正弦波の合成で表したときの最も低い周波数成分の周波数を意味する．基本波とも言う．音楽では音声波形に直接観測される周期(基本周期)の逆数のことである．基本周波数は声帯の振動頻度(ピッチ)によって決まる．簡単に解釈するとピッチの物理量である[4].

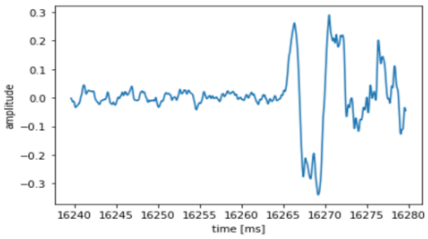

図 2: 基本周波数による抽出した音声特徴量

基本周波数は鼻歌検索、音楽情報検索、楽曲間類似度の推定における応用されている．

4 音響データの種類

音響データの形式は多くの種類がある．デジタルの音声ファイルの仕組みはアナログ信号をデジタルデータに変換して保存される．変換の方法を詳しく解説すると、アナログの波形をデジタルに変換するには、一定の間隔で音の波形を測定する「標本化」という作業が必要になり、その標本化の周期のことを、サンプリング周波数といい．サンプリング周波数は、1 秒間にサンプリングする回数を表す．つまり、1 秒間に何回音を測定して記録するかを表した単位で、値が高いほど原音に忠実になり、高音質となる．また、音質に影響を与えるのはサンプリングレートだけではなく、音の強弱を表す「サンプリングビット」や「チャンネル数」なども影響する．いずれも値が高いほど高音質になる。そして、もう一つ音質に大きく影響を与えるのが、ビットレートである．ビットレートは、1 秒あたりに記録するデータ量を表す[5]．つまり、ビットレートが大きいほど多くの情報を格納できるということで、高音質になる．

これらが、音響のアナログ信号をデジタル化したデータに含まれる情報になる．デジタル化した音響データを、任意のファイル形式に変換して保存することで、その形式に対応するメディア機器やアプリケーションソフトで利用することができるようになる．音響データの保存形式のことを、ファイルフォーマットと言い．ファイルフォーマットは、拡張子が付いたファイル形式のことである．「非圧縮」「非可逆圧縮」「可逆圧縮」の 3 種類のフォーマットに分類することができる[6].

非圧縮は原音のまま記録したフォーマットである、主に AIFF、WAV がある． 圧縮していないため、音質が高くデータ容量が非常に大きい．音響データを原音のまま記録しているため、鼻歌検索のための音声特徴量抽出は主に非圧縮フォーマットを使って実行している．非可逆圧縮は音楽データ大幅に圧縮してパソコンに保存することである、主に MP3、AAC、ATRAC、WMA がある[6]．容量を削減するが音質が下がる．可逆圧縮は非圧縮と非可逆圧縮の間の位置にしている．圧縮時に発生するデータの劣化を最小限に抑え、非圧縮へと再変換可能な形式として開発された．主に AAL、TAK、FLAC がある[6].

まとめ

鼻歌検索における音声特徴量抽出は重要な部分である．音声特徴量はいろいろな種類がある．そのうち、鼻歌検索のための特徴量は主に MFCC、log-mel spectrum、基本周波数がある． 本研究ではこれらの音声特徴量抽出の手法を考査し、実験を行っている．

謝辞

本研究を進めるにあたり，ご助言を頂いた方々に深く御礼申し上げます．

参考文献

[1] 角尾 衣未留, 井上 晃, 西口 正之,「鼻歌検索システムのための楽曲からのボーカルメロディ推定」, Vol.2013-MUS-100 No.18
[2] PowerPoint プレゼンテーション - 音響学入門ペディア http://abcpedia.acoustics.jp/acoustic_feature_2.pdf
[3] 機械学習のための音声の特徴量ざっくりメモ (Librosa ,numpy) https://qiita.com/yutalfa/items/dbd172138db60d461a56
[4] 日本音響学会 - 音のなんでもコーナー https://acoustics.jp/qanda/answer/101.html
[5] 音声ファイル - 主なファイル形式と特徴 https://www.yamanjo.net/knowledge/others_12.html
[6] 音楽ファイル形式の種類と特長は？ https://www.sony.jp/support/walkman/dialogue/010/