

# 音響により検索する音響データベースの試み

7J-7

村松 太一 齊 海 橋本 周司

早稲田大学理工学部

## 1. はじめに

音響を扱いやすくするために、最近では音の内容を考慮した検索システム[1]や音の編集機能を備えた音検索システム[2]が提案されている。我々は以前、音響データを分析して抽出した物理的な特徴を用いて、入力された音響に類似度の高い音響データをデータベースから検索する手法[3][4][5]を提案しているが、この手法では音響データの特徴パラメータが少なかったため、検索の精度が必ずしも良くはなく、ユーザーの要求に見合った検索結果が得られるとは限らなかった。そこで、検索時に用いる特徴の選定を検討し直し、短時間スペクトルから特徴パラメータを抽出して検索を行なう音響データベースの検討を行った。

## 2. システム構成

試作した音響データベースシステムの構成を図1に示す。このシステムでは、音響データから短時間区間ごとの周波数スペクトルを計算しその中のいくつかの特徴パラメータを用いて索引を作る。更にこれらの特徴を索引キーとしてデータベースを構築する。ユーザーがシステムに検索キーとする音を入力すると、同様な方法で特徴パラメータを抽出する。検索は類似検索とし入力音に類似度の大きい音響データをデータベースの中から検索しユーザーに提示する。ユーザーは検索されたいくつかの音響データから目的の音を見つけることができる。なお、検索された結果が目的の音を含まない、または目的の音との類似度が小さい場合は、新たな入力音または検索結果を検索キーとして更に検索を繰り返すことができる。また、このシステムは音響データを加工する機能も有するため、データベース中に目的の音に近い音響データがなくとも、音響データを加工して目的の音に近似させることができる。

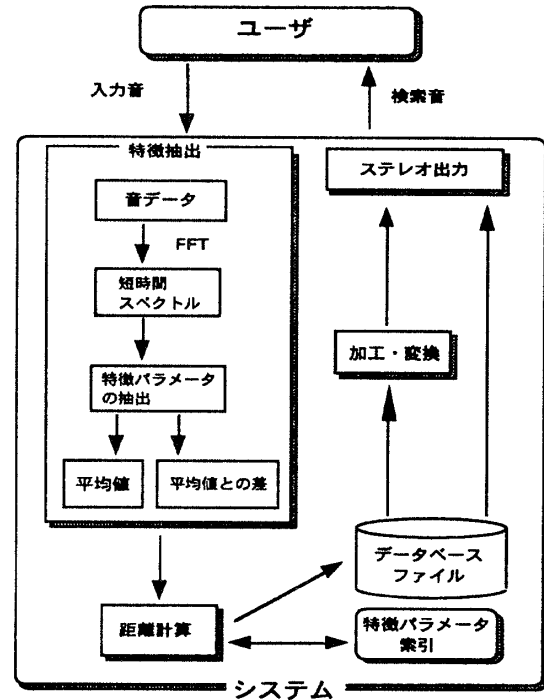


図1. システム構成

## 3. 音響データからの特徴パラメータの抽出

音響データはサンプリングレート 44.1kHz で 16bit 量子化された 65536個のデジタルデータとする。従って、1つの音響データは約1.5秒である。まず、音響データの中で音を構成している部分の時間的な中心点を音響データの中心点に合わせるよう調整する。そして音響データからFFTを用いて短時間区間ごとの周波数スペクトルを抽出する。本システムでは1つの音響データについて128の周波数スペクトルを抽出した。次に、特徴パラメータとして各周波数スペクトルについて図2に示すように、ケプストラム法により基本周波数 $f_0$ を抽出し、周波数スペクトルを8区間に分割しそれぞれの区間での平均レベル $l_1, \dots, l_8$ 、そして基本周波数及びその倍音のパワーと総信号パワーとの比である周期性特徴 $h$ を抽出する。更にそれぞれのパラメータにつきデータ全体の平均を求める $(f_{ave}, l_{ave1}, \dots, l_{ave8}, h_{ave})$ 。次に、各時間ごとに各パラメータと平均との差を計算しそれをパラメータとする $(f_{total}, l_{total1}, \dots, l_{total8}, h_{total})$ 。従って、累計的に使われる特徴ベクトルは1290次元となる。

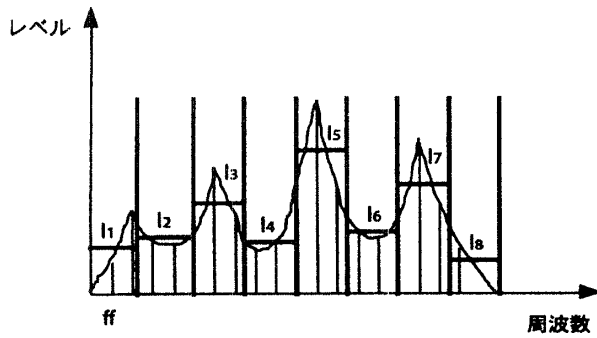


図2. 周波数スペクトルからの  
特徴パラメータ抽出

#### 4. 特徴パラメータによる類似検索

検索は各音の特徴パラメータについて距離を計算し、その値が小さいものから順に選び検索結果とする。また、距離を計算する際に各パラメータの重要性を比で表す検索係数を導入する。検索を繰り返すとこの検索係数が更新され、ユーザーがどのパラメータを重視するかが次第に導き出されていく。距離の計算には次式を用いている。

$$E_i^2 = \sum_n a_n \left( \frac{x_m - x_{in}}{x_m} \right)^2 + \sum_m \sum_i a_m \left( \frac{y_{mt} - y_{imt}}{y_{mt}} \right)^2$$

$n$ :  $ff_{ave}, l_{ave1}, \dots, l_{ave8}, h_{ave}$

$m$ :  $ff_{total}, l_{total1}, \dots, l_{total8}, h_{total}$

$x$ : 特徴パラメータの平均値

$y$ : 特徴パラメータ

ここで、 $r$ は検索キーとする音データ、 $i$ はデータベース中の各音響データの番号である。また、検索係数の更新には、次式を用いる。

$$a'_k = a_k^{x-1} + \left( \frac{1}{\frac{p_{kr} - p_{ks}}{p_{kr}} + 1} \right)$$

$$a_k^x = \sum_j a'_j$$

$k, j$ :  $ff_{ave}, l_{ave1}, \dots, l_{ave8}, h_{ave}, ff_{total}, l_{total1}, \dots, l_{total8}, h_{total}$

$p_{kr}, p_{ks}$ はそれぞれ検索キー、ユーザーの選択した音の各特徴パラメータである。また、次の検索には更新された検索係数を用いる。

#### 5. 検索結果の提示方法

検索結果としていくつかの音響データを提示する場合、文字や画像などのように画面に同時にいくつも表示できるわけではない。そこで本システムでは前回と同様に、ステレオ出力を利用し、複数個の音を同時に流してもそれぞれが異なる位置から聞こえるようにした。そのため、ユーザーは複数個の音を同時に聞き分けたり、それぞれの検索結果の比較、検討を容易に行うことができる。

#### 6. 音響データの加工・変換

本システムはデータベースの検索機能だけでなく音響データを加工する機能も有し、データベース中にある音またはユーザーから与えられた音に対して、その音響データの特徴パラメータをユーザーが調整して元のデータを変換することができる。現段階では、基本周波数の調節や複数音の混合による新しい音の生成などの簡単な処理を行っている。

#### 7. 終わりに

音響をキーとして音響データを検索するデータベースについて、短時間スペクトルを考慮することによって検索精度の向上を図った。提案手法により、個々の音響データの音程や音色の時間的な変化をより反映した特徴パラメータを使用しての検索が実現された。しかし、各音響データについて特徴パラメータを増やすと、検索の精度は向上するものの特徴抽出の計算量が多くなるため検索に時間がかかってしまう。(現システムで約30秒)は音の細かい特徴を利用し且つ検索時間を短縮させるには検索方法の改善が必要である。ここでは音の細かい特徴を重視した検索に重点を置いたが、さらに検索時間の短縮やデータの大きさに依存しない検索方法等を考えたい。

#### 参考文献

- [1] T. Blum, D. Keislar, J. Wheaton and E. Wold: "Audio Analysis for Content-Based Retrieval" Proc. ICMC, 1995
- [2] 和気早苗 旭敏之: "直観的な音データ検索/編集システムの開発" 情報処理学会、情報メディア、29-2、1997
- [3] 斉海 村松太一 橋本周司: "音響データベースのマルチモーダル検索の試み -音響によるデータ検索-" 情報処理学会、音楽情報科学、16-9、1996
- [4] S. Hashimoto, H. Qi and D. Chang: "Sound Database Retrieved by Sound" Proc. ICMC, 1995
- [5] T. Muramatsu, H. Qi, S. Hashimoto: "Sound Database System for Multimodal Data Retrieval -Data Retrieval by Sound-" Proc. International Conference on Computer Music & Music Science, 1996