

電子音色辞書における音色探索システムおよび 3次元音色表示インタフェース

吉岡 靖博 石塚 慎也 斎藤 佳紀 高橋 奈緒子 小坂 直敏

東京電機大学
〒101-8457 東京都千代田区神田錦町 2-2

あらまし

筆者らは、マルチメディアコンテンツ制作に、様々な音色を探索あるいは合成するためのシステム「電子音色辞書」を構築している。本稿では、登録された様々な音色の音色記号と、それから連想されるキーワードや分類から特定の音色を検索するキーワード検索および膨大な音色を3D上に可視化して配置し、類似した音色をインタラクティブに探査する「3次元音色表示インタフェース」について述べる。

The 3D Graphic Interface and The Sound Browsing System for An Electronic Timbre Dictionary

Yasuhiro YOSHIOKA, Shinya ISHITSUKA, Yoshinori SAITO,
Naoko TAKAHASHI, Naotoshi OSAKA

Tokyo Denki University
2-2 Kanda-nishikicho, Chiyoda-ku, Tokyo, 101-8457 Japan

Abstract: In music composition or multimedia-content creation, users need to quickly distinguish the timbre of each sound from the others in a large sound database in order to complete their work efficiently. In this process described in this paper, the Mel frequency cepstral coefficients of a sound object are derived and a principal component analysis for the obtained features is conducted. A sound object is defined as a particle and denoted as a spherical object. Therefore, it becomes easy to search among neighboring sounds in the three-dimensional timbre space by rotating, moving and zooming/widening operations. This user interface is embedded in the sound browsing part of a synthesis system, the "Electronic Timbre Dictionary," and the results are reported.

1. はじめに

近年、コンピュータの普及とともに環境音や効果音を用いたマルチメディアコンテンツ制作が個人でも容易に出来るようになり、関連ツールのユーザも増え続けている。これらの音の音色はコンピュータ音楽には必須の要素であり、音色をユーザの意図に添って検索、合成、加工等を行える機能を有するツールが必要となる。

筆者らは、マルチメディアコンテンツ制作者への音色探索の支援ツールとして、環境音、楽音などあらゆる音を対象とした音色記号からの音検索及び音合成を行うシステム「電子音色辞書」を構築している。これまで、シス

テムの全体構想を提案した[1]。

一方、これまで、音色を可視化し、これをもとに音色の検索を行う方法がいくつか提案されてきた[2] [3]。多数の音色を所持する電子音色辞書では、ユーザとのインタラクティブな対話から検索を行うことが望ましい。

本稿では、まず電子音色辞書の基本的機能を実装した結果について述べ、さらにユーザがイメージする擬音語やキーワードから検索を行う文字列検索について述べる。

また、音声、楽音、環境音、騒音など、音一般を対象にしたデータベースから、それらのスペクトル情報を三次元の距離空間上に表示し、音色の近傍探査など、従来法よりイン

タラクティブで効率的な探査や検索を行うための GUI (Graphical User Interface) として、音色表示インタフェースを実装した結果について報告する。これにより検索、合成など、音色に対する操作性の向上が期待できる。

2. 音色記号

2.1 音色記号の概要

膨大な環境音データベースからユーザが望む音色の検索は困難な問題である。これを解決するため、音色を記号で表現する音色記号[4]を提案してきた。

音色記号は音色を記号化するもので、粒度の大きな順に、巨視的音色、擬音語、微視的音色の三種類に分類して定義する。

説明の便宜上、以下、擬音語から述べる。

1) 擬音語 音色を言語化したものとして定義する。言語としての擬音語は、必ずしも音を模倣したものに限定されず、擬態語と渾然一体となっている。例えば、小川の音色を「さらさら」と定義することは、音の響きをテキスト化したものだけでも言い切れない。音の持つ響きとその情景とが適合するよう選ばれる側面もある。ここでは、具体的な音ファイルに対して、純粋に聞き手の母語に照らしてどのような音韻に聞き取れるか、という純粋に音響知覚的側面を表す。

2) 微視的音色 音色を音声の音韻表記と同等に記述しようとするもので、国際音声記号である IPA(International Phonetic Alphabet)に準ずる記法として定義する。これにより、擬音語の特性である母国語への依存が軽減される。また、新たなユーザにとって全く新しい記号体系よりもなじみ易いと思われる。

3) 巨視的音色 擬音語や微視的音色を複数含み、一つの記号として表現する音色を巨視的音色と呼ぶ。これは、音色を生成的要因で分類するのとは対称的に、音色が類似したものを分類し、包括化したものである。

本検討では、音色記号として、音韻列で表される擬音語と巨視的音色の 2 種類のみを用いている。

表 1 に水音の分類と音色記号を付与した例を示す。巨視的音色では、冒頭の音韻が共通なものをもとめてある。

表 1. 水音の分類と擬音語の例

区分	生成要因			音色記号	
	大	中	小	巨視的音色	擬音語
自然音	水	雨	小雨	s#	a:サー, b:シヤー
			豪雨	z#	a:ザー
		川	せせらぎ	chy*	a:チヨロチヨロ
			急流	d#	a:ドー, b:ドドド

2.2 音色の分類と階層化

ユーザが音色を管理するため、巨視的音色や擬音語を用いた音色記号のみでは、音色の種類が分かりにくい。

そこで音色を生成的、発生的要因から分類を行い、これに基づいた階層化を行った。これにより、ユーザは階層を見れば今選択している音がどのような種類の音色なのかが把握しやすくなる。しかし、階層が複雑になるとユーザへの利便性が減少してしまうため、今回の検討では、4階層までの分類、階層化を行った。

図 1 に分類をフォルダとして階層的に表示した例を示す。まず、自然発生した環境音か人工的に生成した楽音かに分類する。次の階層では、環境音では自然音、交通、生物のいずれかに大分類する。楽音では楽器の種類で分類する。

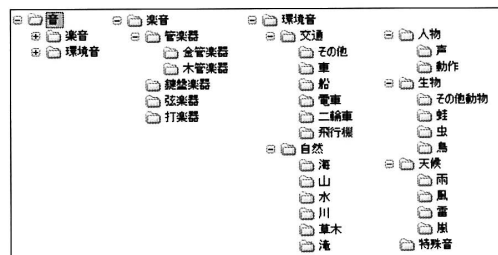


図 1. 音色の階層化例

2.3 音色記号データベースの構築

音色記号を電子音色辞書のアノテーションとして使用するため、XML (Extensible Markup Language) を用いた音色記号 XML を設計する。この XML に音色の id 番号、音色の階層、音波形のファイルパス及びファイルの再生時間、キーワード、コメント、更新日時を属性として付加し、音色記号データベースとして構築する。XML を用いることで、ネットワーク上でのシステム運用や、他のツールやプログラミング言語で容易に利用が可能となる。また、音色に対して、キーワードや特徴を付与するアノテーションを行うことができ、それらを用いた検索も可能となる。

図2に音色記号XMLの一例を示す。この図では、timbre タグにより音色を定義する。また、timbre タグ内の各タグで、前述した音色の様々な属性を表している。

```
<timbre id="1">
  <timbrename>ハッホー</timbrename>
  <tag>ha#</tag>
  <category>環境音/生物/鳥類</category>
  <filepath>animal/cacco01.wav</filepath>
  <time>2</time>
  <keyword>環境音、生物、鳥類</keyword>
  <comment>郭公の鳴き声</comment>
  <lastmodified>2008/12/10 12:31</lastmodified>
</timbre>
```

図2. 音色記号XMLの一部

3. 電子音色辞書

3.1 電子音色辞書の概要

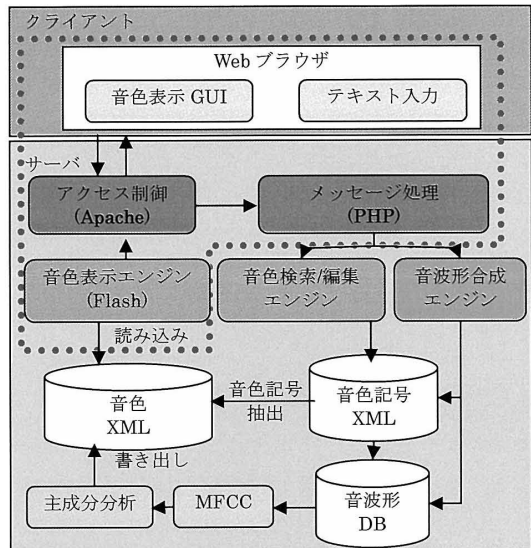
我々は、電子音色辞書を音色の wikipedia と捉えている。多くのユーザがネットワーク上で共有し、音色記号の追加、編集を行うことにより音色記号が安定に収束すると仮定しているため、音色記号と音色のデータベースをそれぞれオンライン上で共有する。

図3にシステムの機能ブロック図を示す。ユーザはWebブラウザ上からGUI操作やテキスト入力により、サーバ内の音色記号データベースにアクセスし、音色記号の検索や登録、及び編集を行う。また、検索された音色記号とその組み合わせから、サーバ内の合成部で音合成や波形接続を行い、新たな音色として登録することができる。

3.2 音色と音色記号の登録

ユーザは環境音のデータと共に音色記号を電子音色辞書にネットワークを通じて登録する。登録する際に、ユーザは音色記号のID番号、擬音語、巨視的音色記号、環境音を生成的視点で階層的に分類した属性、ファイルパス、キーワード、及び内容記述であるコメントを登録する。

図4に登録画面例を示す。図の入力窓にそれぞれのパラメータを記述し、登録ボタンを押すことにより、音色記号データベースに登録する。



*DB: データベース ●●●: インタフェース部

図3. 電子音色辞書のシステム構成図

音色登録	
ID	120
擬音語	ビチャ
巨視的音色	p#
属性	自然音/水/滴
ファイル	water/water105.wav
時間	1 sec
キーワード	自然音/水/滴 p#
コメント	滴が落ちて、石に当たる音を収録 write:saito
登録	

図4. 音色登録画面

3.3 音色記号の一覧表示

電子音色辞書に登録された音色記号をユーザは一覧表示によって確認できる。

図5に電子音色辞書に登録された音色記号の一覧を示す。ユーザは目的の音色記号のグリッドをマウスでクリックすることにより再生され、音色を確認することができる。



図5. 音色記号一覧表示画面

3.4 音色記号の編集

電子音色辞書に登録された音色記号をユーザが任意に確認、変更、削除できるよう、音色記号の編集機能を実装した。

図6に音色記号の編集画面例を示す。編集画面では、ユーザは音色記号の一覧から登録した環境音や環境音の音色記号及びその登録内容を編集できる。

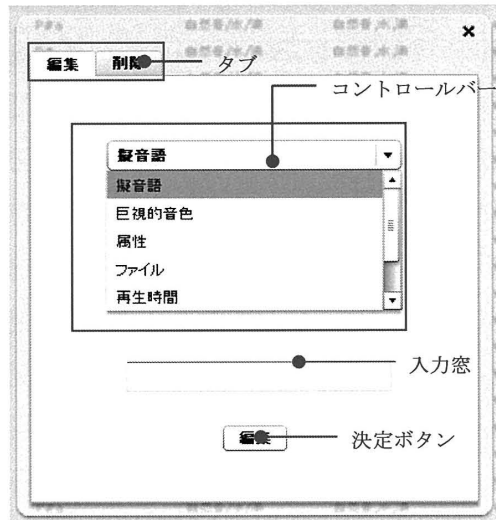


図6. 音色記号編集画面

3.5 音色記号の検索

電子音色辞書が保持する音色記号データベースの一覧からユーザが望む音を再生しながら音色の探索を行うのは非常に困難である。そこで、いかにユーザが持つ音のイメージにそった検索を行えるかが重要である。ここでは、インターネット検索エンジンなどで使用され、ユーザに馴染みやすい検索方法である文字列検索機能を実装した。

文字列検索では、音色記号データベースの擬音語及び、キーワードを対象に、入力した文字列から音色記号の全文検索を行う。

図7に音色記号の検索結果画面を示す。この図では、検索窓に「ポチャ」と入力し、「ポチャ」を含む音色記号の結果を表示している。

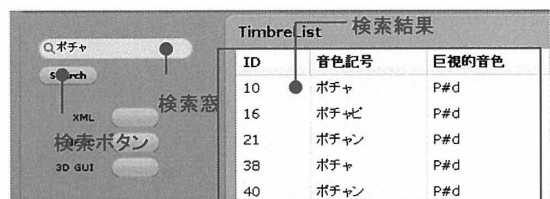


図7. 音色記号の検索結果画面

4. 音色表示インターフェース

4.1 従来の音色の可視化方法

音のデータベースをより効率的に扱うための、音のブラウジング方法として、Real-time corpus-based concatenative synthesis (CBCS)[5],などが知られている。これらは、音楽制作を目的として、大量の音データベースの中から、音の探索や検索を行い、再生および加工などを行うために音データを可視化している。CBCSは、1)2次元上の表示、2)個々の音を点で表現、3)マウスで指示した部分の瞬時再生は可能だが、視点の実時間変更は不可能、などの特徴を持つ。

4.2 音表示と音操作に必要な機能

音色の知覚特性を考慮し、音データの視覚的な表象(音オブジェクト)に以下の特徴を付与した。

- 1.一つの音(一まとまりの音,音脈)に対応した一つの音オブジェクトを独立して存在させる。
- 2.音色とは、個々の音の高さと音の大きさを除く音の総合的印象である。音オブジェクトを音色距離空間上で表現し、聴知覚と対応できるようにする。
- 3.音の探索、加工などの操作を直感と結びつけて効率的に行うために、音オブジェクトに対する操作とその実行・結果表示がリアルタイムで行えるようにする。
- 4.音オブジェクトに対する基本的な操作として、所望の音の再生を最優先とする。

以上の考察から、一つの音脈に対し一つの粒子を対応させ、3次元上に球体で表示する。この表示方法は、天体の表示、分子、原子などの物質モデル、化学式の表示、生体モデルなどのミクロなモデルの表示などで巨視的な視野から微視的な視野まであらゆるスケールでのモデル化の例があり、我々にはなじみやすい。

音の大きさ、音の高さ、音のテクスチャには、それぞれ、粒子の大きさ、粒子の色、粒子の形や表面形状を対応させることができる。

2,3,4では、音オブジェクトを3次元上に配置することにより、音色知覚の直感と対応させ、この空間上で音の見通しをよくし、どの音を対象に、どのような操作を行いたいかをユーザに明確にさせる。

4.を満たすために、マウス操作で、空間移動と視野の拡大/縮小、視点の回転など、アフィン変換操作を天体の間を超高速で動くロケットのように、リアルタイムで行えるようにした。

4.3 音色表示インタフェースの概要

ユーザが望む音色の音色記号や擬音語をユーザ自身が明確な場合、文字列検索により検索できる。しかし、類似した音色や近傍音を探索していくことは、文字列検索機能では難しい。

また、ユーザがイメージする音色記号や擬音語が電子音色辞書の保持する音色記号と一致するとはかぎらない。よって、膨大な音色を保持する電子音色辞書においては、ユーザとのインタラクティブな対話から検索を行うことが望ましい。

そこで検索のための GUI 上に音色記号群を3次元空間上にオブジェクトとして配置し、マウスカーソルで選択することでユーザがインタラクティブに音色を検索し試聴できる、音色表示インタフェースを実装する。これを利用することで、音への操作を近傍音色の空間的な配置の中で、音色の位置付けを確認しながら行うことが可能となる。

4.4 音色の表示方法と3次元データへの変換

各音色に対し、MFCC (Mel-Frequency Cepstrum Coefficient) [6]を用いて、24次元の特徴量を抽出する。音の高さに対して人間の知覚は、低い周波数ほど分解能が高く、高い周波数ほど分解能が低くなっており、それらはほぼ対数に近い非線形の特性を持っているが、これらの聴覚特性を反映させたスペクトル分析方法がMFCCである。この特徴量を表示可能とするため、主成分分析により、3次元に集約する。

4.5 音色XMLの設計

電子音色辞書が持つ各音色を、音オブジェクトとして、3次元空間上に表示したい。これを可能とするため、音オブジェクトをXMLとして記述した、音色XMLを設計する。音色XMLでは、音色記号データベースである音色記号XMLから抽出した各音色に対し、前節で述べた3次元データを付加して作成する。

図8に音色XMLの一部を示す。この図では、`timbreobject`タグにより3Dオブジェクトを作成している。`timbreobject`タグ内の`timbre`タグの要素で音色記号の一種である巨視的音色を表す。その中の`id`属性により音色記号XMLと対応する。同様に`filepath`タグにより音波形データへのパスを表す。これにより音色の利用が可能となる。また、`position`タグの属性に、主成分分析により求めた第1,2,3主成分を空間上に配置するため、それぞれ`x,y,z`に表記する。

```

- <timbreobject id="1">
  <timbre id="6">S#</timbre>
  <filepath>./snd/s001.wav</filepath>
  <position x="300" y="200" z="300" />
</timbreobject>
- <timbreobject id="10">
  <timbre id="1">S#a</timbre>
  <filepath>./snd/s005.wav</filepath>
  <position x="350" y="400" z="350" />
</timbreobject>

```

図 8.音色記号 XML の一部

4.6 音オブジェクトの 3D 表示

4.5 節で述べた音色 XML を読み込み、音色記号を 3D オブジェクトとして 3次元空間上に粒子として表示する。また、簡易なアニメーションとして、音色記号を表示する。

音オブジェクト表示の実装には、Web 上のブラウジングとするため Flash を用いた。これにより 3D 空間上においてマウス操作およびキーボード操作を行うことができる。また自由に視点を移動、回転することで類似した音色をインタラクティブに探索できる。

4.7 音色表示インタフェースの試用実験

図 9 に音のブラウジングの画面例を示す。この図では巨視的な音色である「S#」とその近傍の音色を表示している。この表示機能により、音色へのアクセスが容易となり、音に対するさまざまな操作をこの音色表示インタフェースを通じて行うことができた。

また、ユーザがより能動的に音色の検索を行えるよう、GUI 上の音色記号群をマウスカーソルでなぞることで連続して試聴できる機能「蛇道試聴」[1]を実装した。同図の上部では「S#」内の音色を左上から右下へマウスカーソルでなぞって連続で試聴を行うことができる。

試用実験の結果、音色表示インタフェースの表示機能により、近傍の音色へのアクセスが容易になることが確認できた。

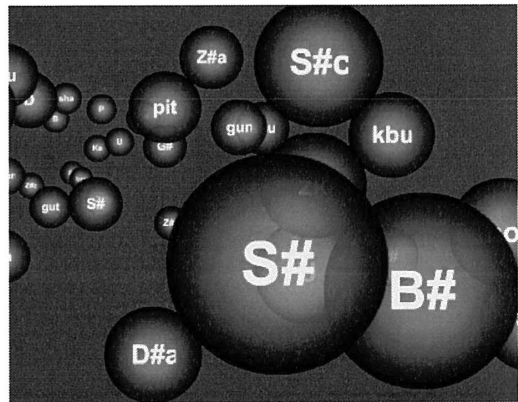


図 9. 音色表示インタフェースの画面例

5. おわりに

本稿では、音色探査のための音色記号を用いた検索システム電子音色辞書の新しい文字列検索機能および、大量の音データを効率的にインタラクティブに探査するための音色表示インタフェースについて必要な機能を考察し、その実装を行った。また、これを電子音色辞書に応用し、試用実験を試みた。今後の課題として、音合成系への応用や近傍音での知覚評価実験を行うことが望まれる。

謝辞 本研究では、科学研究費補助金(課題番号 20520134)の補助を受けて行われた。

参考文献

- [1] 小林, 小坂, “音色記号による環境音検索のための電子音色辞書編集システムの構築,” 音講論(春), 449-452, 2008.
- [2] M. Fernstom, and P. Cook, “Enhancing sonic browsing using audio information retrieval,” Proc.ICAD02, kyoto, 2002.
- [3] G. Tzanetakis, and P. Cook, “3D Graphics tools for sound collections,” DAFX-00, verona, Dec., 2000.
- [4] 小坂, “楽曲制作のための音色理論の構築にむけて,” 音講論(春), 3-7-15, 2005.
- [5] 小坂, “楽曲制作のための音色理論の構築にむけて,” 音講論(春), 3-7-15, 2005.
- [6] D. Schwarz, S. Cadars, and N. Schnell, “What next? Continuation in Real-time Corpus-Based Concatenative Synthesis,” Proc. ICMC 2008, Belfast, 2008.
- [7] M. Hunt et al., “Experiments in syllable-based recognition of continuous speech,” in Proc.1996ICASSP, pp.880-883, 1980.