

音色記号による音合成のための電子音色辞書編集システムの検討

小林 洋平 吉田 誠一 後藤 祐子 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

E-Mail: yohei@srl.im.dendai.ac.jp, osaka@im.dendai.ac.jp

あらし

本報告では、音色記号を用いた音合成を行う電子音色辞書編集システムについて検討する。それまで試行錯誤によって合成されている効果音も、音楽合成、音声合成と同様の自動合成の枠組みで合成が可能であることを提案する。まず、音色記号の粒度とその特徴について考察し、それに基づいて音色の分類、記号定義を行う。それらを蓄積、管理する電子音色辞書と、それから記号を検索、編集し、それらを組み合わせた記号列を入力とした音合成を可能とするシステムを構築する。また、システムを利用する際、音色記号の検索に有用な二つのGUIについて検討する。

A study on electronic timbre dictionary editing system for sound synthesis by timbre script.

Yohei KOBAYASHI, Seiichi YOSHIDA, Yuko GOTO, Naotoshi OSAKA

Tokyo Denki University

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

E-mail: yohei.yoshida.goto@srl.im.dendai.ac.jp, osaka@im.dendai.ac.jp

Abstract

This paper describes a new notation based electronic timbre dictionary system. First of all, the definition of the timbre notation in terms of the granularity of timbre notation is given. Then, the concept and the outline of the system are introduced, including sound engines and two graphical user interface types are introduced.

1. はじめに

近年、映画やゲームをはじめとするマルチメディアコンテンツにおいて、音楽、音声、効果音などの音コンテンツは様々な場面で使用され、これらに用いられる音合成技術やシステムは数多く存在する。この中で、音楽合成は一般にシーケンサと呼ばれるシステムに譜面情報を入力する。システム内部ではこれらから音高列と音長列のシンボル列を根幹とし、音の強弱や演奏表情の情報を加え合成する。また、MIDI (Musical Instrument Digital Interface)信号を利用し、異なる電子楽器間で記号列、および演奏表情の情報通信を行った合成も可能である。音声合成の分野では、素片編集型合成[1]を例にとると、テキストデータを入力として、音素列やピッチ列、音長列を算出し、それらを用いて音素片辞書から音素片を取り出し、これらを接続して合成音声を作る。

これに対し効果音の合成は、一般に収録された音データ、またはモデル音にユーザが対話的にエ

フェクトを加えて目標とする音色に近づけるような音合成を行う。音楽合成と音声合成のように記号系列に基づく自動合成の枠組みはなく、試行錯誤による合成手法が一般的である。この方法は、一つ一つ丹念に高品質の音を作り上げるためには必要な機能だが、1)高速かつ大量の音合成、2)使用した音色および合成手法の表現、記録、伝達などの機能、3)オフィスツールのように、非専門家が簡便に音を使用するための音合成を達成するために、音色を記号化することが必要である。

以上を考慮し、本研究は音色記号を用いた音合成を行うシステムの構築を目標としている。まず、音の知覚的な単位となる音色記号を粒度別に分類し、その特徴について考察する。それらに基づき、対象とする音色を分類し、記号や擬音語を定義[2]する。また、音色の数に比例して増大する音色記号を蓄積、管理する電子音色辞書について検討する。また、電子音色辞書から音色記号を効率よく探査、編集を行うための機能と、それらを組み合わせた記号系列や時間長列などを入力と

し、音合成を行う機能を持つ電子音色辞書編集システムについて検討する。これにより、効果音や新しい音楽に対しても音楽／音声と同様の枠組みでの自動合成が可能となる。また、これらの記号系列をスクリプトとして記述することにより、合成手法の表現、記録、および伝達を正確に行える。

2. 音色記号の設計

2.1 音色記号とその粒度

音色空間は多次元的であり、客観的な音色記号の定義は存在しないと考えられる。そこで、これを客観化することは断念し、ユーザが定義したとき、これを用いて音合成するというメタなレベルでの手法を考案する。ここでは音色記号を音色素と呼び、ユーザは記号をそのレベル（粒度）とともに主観的に定義できるものとする。音声合成は C/V などの音韻が一般的である。音楽合成では、二つの波形素片を組み合わせて合成する手法は **diphone**[3]として知られている。しかし、音二つ以上を有機的に組み合わせるユーザの意図する合成音を作る枠組みにはなっていない。そこで、音の知覚的な単位である音色素を設定し、これらの複合体として音を表現する方式を表 1 に示す。表 1 の 1,2 の音色素は知覚的な単位であるが、同時に物理的単位でもあり、粒度も小さい。これらは、正弦波モデル[4]、SMS(Spectral Modeling Synthesis)[5]としても知られ、音声符号化のための基本方式でもある。5,6 は言語音を用いる音韻を一般の音に適用しようとするものである。擬音の構成要素が音韻であるため、粒度をそれぞれ大、中と区別した。3,4 は、必ずしも言語音の音韻と対応させず、音韻外の記号で表記する。また、微視的音色は記号の時間が音韻よりも短いもので定義し、巨視的音色記号は、微視的音色記号の要素として表現する、あるいはひとまとまりの音色を一つの記号で表現する場合をいう。一方、5,7 は音楽音をスペクトル歪みに基づき音韻列で表したり、逆に音声を楽音列で表す研究[6]である。これは音色を考えるツールとなるが、得られる合成音は粒度が大きすぎ、表現したい原音との差が著しい。むしろ、アミューズメントなど別の目的での使用が考えられる。本研究で用いる記号の粒度はこれらの中間に行くもので、微視的音色、巨視的音色、擬音語、音韻、音素などとする。

2.2 音色記号の定義

2.1 に基づき、対象とする音を電子音や自然音などの生成状況、孤立発声や連続発声の発声状況、

表 1 音色素による音色記述の分類

No.	音色素	粒度	測度	分析法
1	正弦波	小	S/N	文献 4
2	正弦波、雑音	小	SD*	文献 5
3	微視的音色	中	SD,その他	本研究
4	巨視的音色	大	SD,その他	本研究
5	擬音語	大	SD	本研究
6	音韻/音素	中	SD	文献 6
7	楽音 (楽器名、音色)	大	SD,知覚歪	文献 6
8	MIDI 信号	大	SD	文献 7

* SD: Spectral Distortion

表 2 自然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	
	滴	→	p#	a:ピチャ, b:ピシヤ, c:ピタ d:ポチャ, e:ポチ, f:ポワン	
	火	焚き火	→	p*	a:/バチバチ
		花火	打ち上げ	h#	a:ヒューン, b:ヒュルルル
			手持ち	d#	a:ドーン, b:ダーン
	風	強風	→	b#	a:ビュオー, b:ビュー
				g#	a:ゴー
木枯らし		→	hy#	a:ヒュオー, b:ヒュルルル	

そしてそれらの知覚に基づいて分類し、記号や擬音語を定義する。表 2 に自然音の分類と擬音語の例を示す。この例では、左から自然現象に基づいて「水」、「火」、「風」など粒度の大きい生成要因に分類し、さらに各要因内で細分化していき、最下層で巨視的音色と擬音語を定義している。

3. 電子音色辞書編集システムの構築

3.1 システムの概要

本研究で対象とする音は数多く存在する。なぜなら、全ての音が効果音となり得るからである。楽器を用いた音も、不協和音などにより、演奏者の意識、無意識に関わらず楽音として聞こえないことがある。また、工事現場などから聞こえる騒音も、ある一定のリズムが付くことで、ドラムやパーカッションのように聞こえることがある。このように、同じ音でも発音のタイミングや音色の組み合わせなどで、聞こえ方が変わってくる。よって、本研究では生成要因に関わらず全ての音色を扱う。

同様に、同じ音でも聞こえ方やその分類は個人や言語圏などで異なり、一意に定まらない。よって、定義された音色記号や擬音語も膨大になる。そのため、音色記号の定義を個人、または数人で

完成させることは困難である。そこで、wikipediaと同様に、音色記号をデータベース化し、オンライン上で共有することで、多くの人が分類や記号定義などの編集、およびそれらを用いた音合成を行える電子音色辞書編集システムを構築する。図1にシステムのブロック図を示す。

本システムは電子音色辞書編集部と音合成部の二つを柱としている。また、ユーザがこのシステムを利用する際、インターフェイスとして編集モードと自動合成モードの二つを設ける。編集モードでは、GUI やコマンドを用いて編集部にアクセスし、電子音色辞書に対して音色記号の検索や編集を行い、その記号を入力として音合成部でサンプル音の合成を行う。自動合成モードでは制作者がスクリプトとして記述した音色記号列を入力とし、同様に音合成を行う。また、オンライン上でこのシステムを運用する際、外部からの利用を管理するネットワーク制御部を設ける。

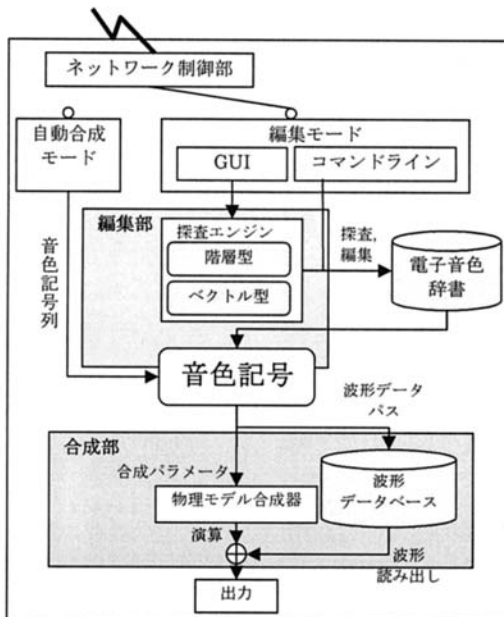


図1. 電子音色辞書編集システムのブロック図

3.2 電子音色辞書編集部

システムを利用する際、膨大かつ未知、あるいは不慣れた音色記号を単純にリストなどで提示するだけでは、制作者の作業効率を下げる。制作者の意図する音色記号を素早く見出すためには、検索を容易にする探索機能が必要である。また、記号の追加定義や対応する音サンプルや合成プログラムの提示、音色の特徴を要因として付加するなどの編集機能も必要である。現在は、コマンドを用いた検索のほか、音色記号の特徴を基に、検索に有用な2つのGUIを検討している。

3.2.1 階層型 GUI

2.2 より、音色記号はその粒度に応じて階層的な構造にすることができる。すなわち、最下層では擬音語とし、その上位ではそれらを包括する抽象的な記号とする。制作者は、どの粒度でも自由に記号を付与でき、この階層構造をGUIによって表現する。制作者が望む音色を明確に表現できず、抽象的な想像のみで留まっているときに、上位の階層から順に辿ることで、所望の音色を見付けられる。逆に、明確な音色の指定はなく、ある程度の粒度の音色を必要とするとき、上位の抽象記号を選択することで、システム側でその階層以下にある音色から自動で選択させることができる。図2に水音の一部の階層を辿った場合のGUIの例を示す。制作者が厳密に音色を指定したいときは、最下層の擬音語を選択し、「滴」や「水音」であれば良いときは、その上位の記号を選択するなど、各階層を自由に行き来し、所望の音色を見付けられる。

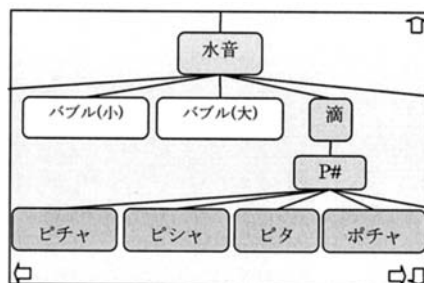


図2. 階層型 GUI の例

3.2.2 ベクトル型 GUI

3.1.1 の階層的な構造とは別に、音色を距離空間上で表現できると、Google Earth と同様の近傍探索、移動など、GUI 探索が可能である。そこで、音色の階層構造とは別に、音色に関する要因を複数設け、それらの数値を要素とするベクトル空間を考える。図3にある水滴音を選択したときのベクトル型 GUI の例を示す。ベクトルのそれぞれの軸に要因を任意に割り当て、最初に選択された音色を中心に、他の音色を表示している。また、各軸に割り当てられた要因は、製作者が自由に変更することができ、その都度候補となる音色を自動で変更、再提示する。制作者は選択している音色が所望の音色と異なるとき、各要因を参考に別の音色を選択し、その音色を中心に再度ベクトル空間を更新する。これを繰り返すことで所望の音色へ素早く到達できる。

3.3 音合成部

編集部で検索し、選択された音色記号や、スクリプトとして記述された音色記号列を入力とし、音合成を行う。図4に各入力からの音合成のブロック図を示す。ユーザは編集部で検索、選択した音色記号や、音色記号列や時間長列、ピッチ列などを記述したスクリプトを合成部に入力する。

音合成部では、辞書の機能を満たすため、できるだけ多種多様な音を合成できることを理想とする。まず、多くの波形データを持ち、記号が入力されたとき、これから音波形の直接的利用、あるいは、加工して接続する素込接続機能を有す。また、物理モデルによりできるだけ多種の楽音あるいは自然音を合成できることを目標としている。また、記号に対応した音合成は一意ではなく、複数の中から選択できるものとする。例えば、クラリネットの音の合成に波形データ、および物理モデルの両方が使えるとき、制作者はそれぞれを比較して良い方、あるいは便利な方を自由に選ぶことができる。

現在、物理モデル合成エンジンとしてScavone, Perry らのSTK (Synthesis ToolKit)[8]、引地らによる笙の物理モデル[9],[10]を実装している。物理モデルによる音合成は、実在の楽器と同様なパラメータの操作が可能であるため、より直感的で簡便な音色のコントロールや、実際の楽音の特徴を有する合成音が可能である。また、楽器の長さやリードの固有各周波数なども制御パラメータとして利用できるため、実在の楽器では表現できない音色の合成も期待できる。現在実装している楽器の一覧を表3に示す。

3.3.1 STK

STK はC++ により実装された音響信号処理と合成アルゴリズムのクラスのオープンソースである。クラリネットやサククスなど、洋楽器を中心とした様々な楽器の物理モデルが用意されている。それらを用いた音合成や、SND, WAV, AIFF などの各種サウンドファイルへの出力が可能である。

3.3.2 笙の物理モデル

STK は洋楽器を中心とした物理モデルであるが、本システムではより多くの種類の楽器を扱う。現在は、邦楽器のひとつである笙の物理モデルを実装している。笙は主にマウスピースと匏(ほう、又はかしら)と呼ばれる空洞部分と 17 本の竹管で構成される、フリーリード楽器である。今後はさらに多くの邦楽器など、STK で扱われていない物理モデルを実装する。

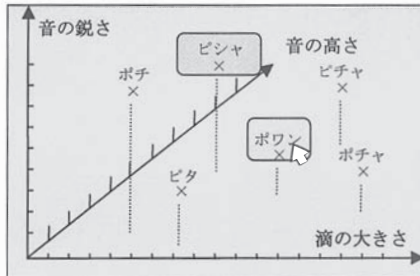


図 3.水滴音のベクトル型 GUI の例

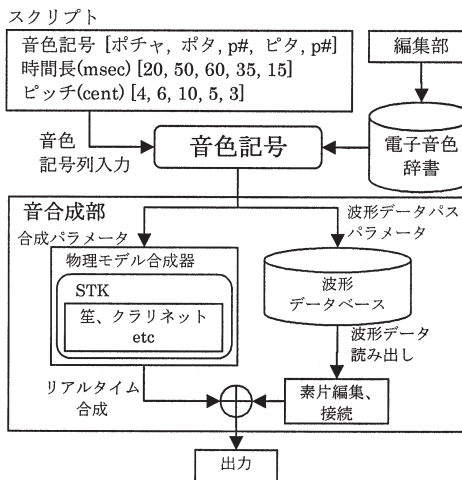


図 4. 各入力からの音合成のブロック図

表3 実装している楽器

洋楽器	クラリネット、擦弦、撥弦、撥弦(硬い弦)、サククス、マンドリン、シタール、BlowHole
和楽器	笙

4. 音色記号による音合成の利用法

1章で述べたが、音色記号による音合成が目標としている機能を再度示す。

- 1)長時間の効果音などを高速かつ大量に合成
- 2)使用音色や合成手法の明確な表現、記録、および伝達
- 3)オフィスツールなどで、非専門家が効率よく音を利用するための視覚的合成

上記の機能を達成するために、電子音色辞書編集システムとは別に、上位レベルのツールを用いた音合成について考える。ここでは、1)、2)の機能を達成するためにシーケンサと、3)の機能を達

成するためにオフィスツールと連携し音色記号の記述や選択から自動合成を行う枠組みについて考える。図5に上位ツールと組み合わせた音合成のブロック図を示す。各ツールは電子音色辞書編集システムの編集モードと自動合成モードを切り替えて利用し、音色記号の検索、記述、および自動合成を行う。

4.1 シーケンサを用いた自動合成

電子音色辞書編集システムでは一つ一つの音色に対して記号の編集や検索、音合成を行える。しかし、マルチメディアコンテンツにおいて、効果音をはじめとする音コンテンツは複数の音色を同時、そして連続して用いて合成されている。

そこで、シーケンサを利用し、音楽合成と同様の枠組みで効果音の自動合成を行う。図6にマルチメディアシーケンサの一つのトラックに音色記号用のトラックを実装した例を示す。制作者は音符の記述と同様に、音色記号を自由に記述することができる。また、各音色記号の時間長やピッチなど、合成に必要なパラメータの編集機能を備える。それらの記号列を、スクリプトとして電子音色辞書編集システムの合成部へ入力し、音合成を行う。これにより、制作者が意図する音を合成するまでの過程を明確に記述することができる。また、音色を一つ一つインタラクティブに合成する必要はなく、高速かつ大量に合成を行うことができる。さらに、音楽における楽譜と同様に、合成手法の記録や他者への伝達、および再現が可能になる。

4.2 オフィスツールでの利用

オフィスツールで音オブジェクトを利用する際、各オブジェクトに対して単一の画像によって表示される。クリップアートなどの画像オブジェクトとは異なり、視覚による取捨選択はできない。また、googleをはじめとする検索エンジンは、所望とする画像のキーワードを入力することで、サムネイルによる一覧表示や、その画像を取得することが可能である。しかし、音に関しては同様の機能は無く、実際に音を聞いて利用するか否かを判断するため、特に非専門家が音オブジェクトを扱う際、作業効率が低下する。

そこで、音色記号を利用し、視覚による音オブジェクトの選択を可能にする。図7にオフィスツールでの利用例を示す。ユーザは右側に表示されている電子音色辞書の2つのGUI、またはコマンドラインを使い、必要とする粒度の音色記号を検索し、メイン画面にドラッグアンドドロップする。その際に音合成を行い、音データとして保存

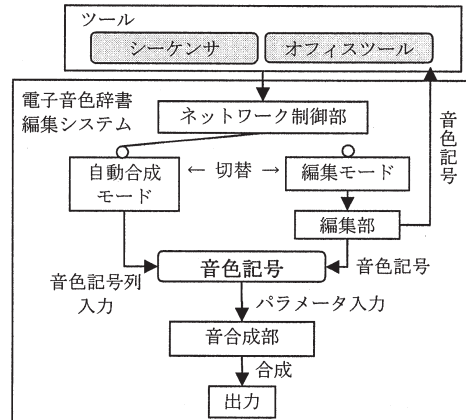


図5. ツールを用いた音合成のブロック図

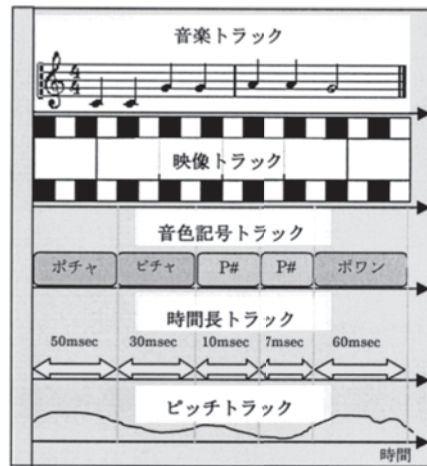


図6. シーケンサを用いた音合成の例

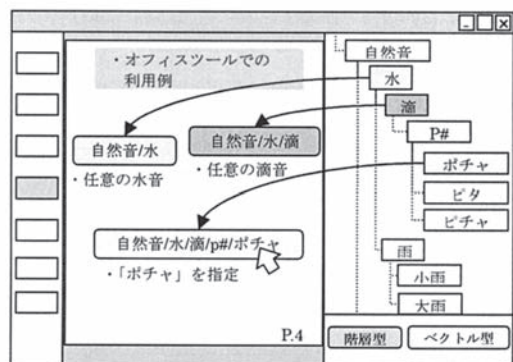


図7. オフィスツールでの利用例

し、記号と対応づける。音色記号を表示することで、画像検索と同様に、視覚による音オブジェクトの取捨選択が可能になり、より短時間で作業できる。

5. おわりに

本報告では、音色記号を用いた電子音色辞書編集システムについて検討した。まず、音の知覚的な単位となる音色記号の粒度とその特徴について考察した。そしてそれらに基づいて実際に音色の分類、記号の定義を行った。それらの音色記号を蓄積、管理する電子音色辞書と、記号の検索や編集、および記号列を入力とした音合成を行うシステムについて検討した。また、検索に有用な階層型、ベクトル型の GUI について検討した。音合成方式に、波形データを連結する素片接続による方式と、STK と笙の物理モデルを用いた合成方式について述べた。また、シーケンサとオフィスツールの二つの上位ツールを用いた利用法について述べた。これにより、それまで試行錯誤によって合成されてきた効果音も、音楽/音声合成と同様の枠組みで行うことができることを提案した。

今後の検討として、音色の分類や記号定義を進めるために、電子音色辞書のオンライン上での共有を行う。初期段階では内部でネットワークを通じて音色記号の定義やシステムの利用、改善を行い、性能と時期を見て外部に公開する。また、より具体的な GUI の構築も行う。

参考文献

- [1] 小坂 直敏, “楽曲制作のための音色理論の構築にむけて”, 日本音響学会講演論文誌, 3-7-15, 2005.
- [2] A. Iida, N. Campbell, “Speech Database Design for a Concatenative Text-to-Speech Synthesis System”, *Int'l Journal of Speech Technology*, Vol.6, Issue 4, pp.379-392, 2003.
- [3] Rodet, P. Depalle, *Synthesis by Rule: LPC Diphones and Calculation of Formant Trajectories*, IEEE-ICASSP, Tampa, FL, March 1985.
- [4] McAulay R. J. McAulay and T. F. Quatieri, “Speech Analysis/ Synthesis Based on a Sinusoidal Representation,” *IEEE Trans. on Acoust., Speech, and Signal Processing*, vol. ASSP-34, No. 4, Aug. 1986.
- [5] Serra, X. and J. Smith, “Spectral modeling Synthesis: A Sound Analysis/Synthesis System based on a Deterministic plus Stochastic Decomposition.” *Computer Music Journal* 14(4): 12-24, 1990.
- [6] Rodrigo Segnini, “Score phonetization and speech derived composition,” *Proc. of ISMA (International Symposium on Musical Acoustics)*, 2-S2-13, Nara April 2004.
- [7] Modegi, T., “Development of MIDI Encoder Tool ‘Auto-F’ for General Time-Based Electric Signals,” *ICAD2002 (The Eighth International Conference on Auditory Display)*, pp158-163, Kyoto, Japan, 2002.
- [8] P. Cook, “Introduction to Physical Modeling”, A K Peters Ltd, *Audio Anecdotes : Tools, Tips, and Techniques for Digital Audio*, pp.179-198, 2003.
- [9] Hikichi, Osaka, and Itakura, “Timedomain simulation of sound production of the sho,” *Acoustical Society of America*, Vol.113, No.2, pp.1092-1101, 2003.
- [10] Hikichi, Osaka, and Itakura, “Sho-So-In,” *Journal of New Music Research*, Vol.33, No.4, pp.355-365, 2004.