

研究報告

電子音色辞書とその創作への応用可能性 THE ELECTRONIC TIMBRE DICTIONARY AND ITS APPLICATION TO CREATION

小坂 直敏

Naotoshi Osaka

東京電機大学 未来科学部

School of Science and Technology for Future Life

Tokyo Denki University

概要

われわれの研究室では音のアーカイブとしてネットワーク上に「電子音色辞書」を構築している。これは、音の Wikipedia と同様に、ユーザはネットワークから自由に音を検索および合成、登録を行うためのシステムである。Wikipedia から類推される機能、あるいは既存システムとの相違は、音の検索のみならず、新たに合成する機能を付加している点、また、音色を擬音語から引き出すように、音のイメージから音を検索する逆引き機能も有す点である。研究として技術的な検討課題は、1) システムとしての全体構成のありかた、2) ユーザが音を理解、把握しやすくするために、どのような GUI(Graphical User Interface) を設計すべきか、また 3) 音のイメージからの逆引きを可能とするための音色記号の定義などである。本稿では、これまでに実装できた部分を紹介し、さらに、音楽やサウンドスケープなど創作の場への応用例を紹介する。

In our laboratory, the “Electronic Timbre Dictionary”, a type of internet sound archive, is being constructed. This can be said to be a Wikipedia for sound. The system enables a user to search, synthesize and register sounds freely on the network. The differences between known systems such as Wikipedia and our system are that 1) not only search functions but synthesis functions are provided, and 2) an inverse look-up of sound from a user’s sound image such as onomatopoeia becomes possible. Technical areas of study include 1) the total system configuration, 2) GUI (Graphical User Interface) design, and 3) timbre symbol definition in order to enable inverse look up. In this paper, already-implemented functions are introduced, centering on the GUI, and examples of its application to the music creation and soundscape are introduced.

1. はじめに

電子音響音楽創作やマルチメディアコンテンツ制作において、音素材は大変重要な位置を占める。用いる音素材は、収録音や合成音のほか、既存の音データベースからの検索によりデータを準備することもある。これらにより、入手した音素材は、さまざまな処理により加工された音が、再び別な場面での一次素材となることもあり、入手後も、効率的な整理ができる方法が望まれる。

筆者らは、このようなニーズに応えるべく、コンピュータ音楽やマルチメディアコンテンツ制作者への音色操作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、音の登録や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」を構築してきた[小林 08, 吉岡 09, Osaka 09]。

電子音色辞書は、音の Wikipedia と同様に、類似したシステムに freesound [mus] がある。これは既に実用化され、現在誰もが利用できる。データ数も膨大となっており、充実した音データベースとなっている。Freesound に搭載されていない機能として、音の合成と、音色からの逆引きがある。この二つの機能を達成し、さらに、より洗練された GUI(Graphical User Interface) についても検討を行っていくのが当電子音色辞書プロジェクトの狙いである。

本稿では、まず、電子音色辞書の基本的機能を中心に、その概要を紹介する。特に音色のイメージから音を逆引きするために導入された音色記号について述べる。実装の詳細については上記文献を参考にされたい。

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[Fernstom 02, Tzanetakis 00]。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。

そこで、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3次元の距離空間上に表示し、音色の近傍探索など、従来法よりインタラクティブで効率的な探索や検索を行うための GUI として、Sound Planet について紹介する。これにより音色探索時に操作性の向上が期待できる。

次にこの GUI を発展させることにより、構築できる新たな創作の枠組みを、音楽とサウンドスケープの両方について考察する。

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探索は困難な問題である。これを解決するため、音色記号[小林 05]を提案してきた。

音色記号は、音色を粒度の大きな順に、巨視的音色、擬音語、微視的音色の三種類に分類して記号を定義するものである。

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、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。すなわち、聴覚的視点での分類である。微視的音色と擬音語については、われわれが発音できる音韻を参照物とし、これで表現することを意図しているが、巨視的音色は必ずしもその意図はない。例えば、専門家のみがわかるエンジン音の異音を A 音と記すなど、共通の印象を与えるものをくくるための目的の記号であり A が音の印象を表しているわけではない。

本稿では、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まとめている。

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:シャー, b:シャー
			豪雨	Z	a:ザー
		川	せせらぎ	chy*	a:チョロチョロ
			急流	d	a:ドー, b:ドドドド

表 1. 水音の分類と擬音語の例

2.2. 音色の分類と階層化

巨視的音色や擬音語を用いた音色記号のみでは、ユーザが音色を管理するためには不便である。例えば収録時のデータは発音体に応じて分類しておく方が实际的であり、聴覚的分类はデータ整理後の区分である。また、音の探索も生成的側面と聴覚的側面の両方からできる方が便利である。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。



図 1. 音色の階層化例

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。

なお、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 を設計した。

揮するであろう。

3. 電子音色辞書

電子音色辞書は、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2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。

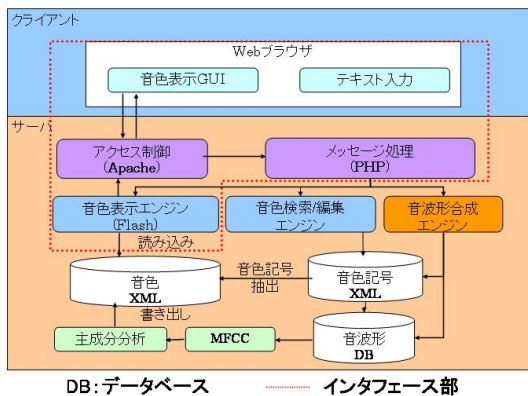


図2. 電子音色辞書のシステム構成図

3.1. 電子音色辞書の応用

本システムは、これまで紹介してきた、freesoundと同様の機能のほか、いくつかの応用が考えられる。1つはパーソナルユースがある。ユーザ自身の創作工程で音の再加工、リミックスなどを経ることにより膨大な中間的音素材が合成される。これらを効率的に格納するためのツールとして応用できる。この場合の音色記号は、客観性は必要なく、まったくパーソナルユースでよいため、記号の収束性についての悲観的な考察は不要となる。

また、もう1つは、専門家で閉じたコミュニティ内での音データベース管理ツールである。車のエンジン音の異音検出、メーカ特定などの目的で、エンジン音の専門家のみで共有できるデータベース、あるいは、内科医の聴診器からの診断音の分類、異常な嚔下音やさ声など、声も含めた生理的な音の中で異常なものを集めた医療関係者用のデータベースなどである。これらはデータが守秘的である、という側面とその音色の共通性が専門家でないとわからない、という特殊性がある。このような場では巨視的音色、というものが多いにその価値を発

4. 音色表示インタフェース

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)[Schwarz 08]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、音の探査や検索を行い、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. 音の探査、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度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が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。そのため、膨大な音色を保持する電子音色辞書においては、ユーザがシステムとインタラクティブに探索することが望ましい。

そこで探索のための 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 Sound Planet と呼ぶ。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)[Hunt 80]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徴を持っているが、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。図 3 に表示画面の例を示す。

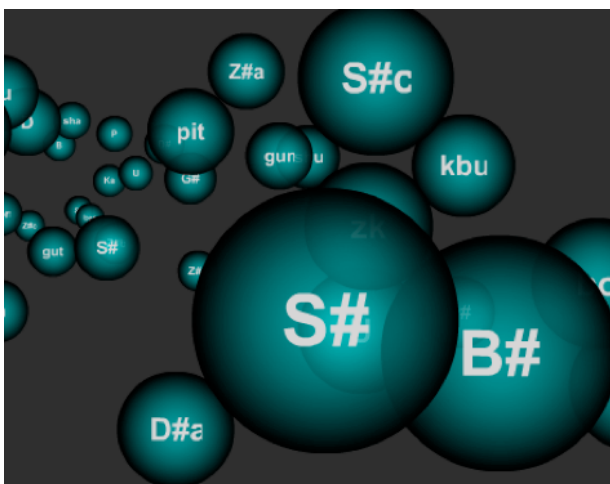


図 3. 音色表示インタフェースの画面例

	要因	現音表示 GUI の表現	発展系
1	対象	一つの音、音脈 ストリーム	複合音 音楽
2	表示法	球、粒子	任意形状/ 波形、スペクトル
3	時変/ 時不変	時不変	時変
4	布置	聴覚と対応の良い 特徴量	任意に定義

表 2. 現状の音表示 GUI とその発展形

4.5. 創作への応用可能性

さて、以上のべたシステムは電子音色辞書として、現在も一般公開を目指して検討と実装を進めている。本稿では、これを作曲システムに応用する発展の可能性について論ずる。

まず、現在のシステムではサウンドオブジェクトを 1 つの粒子で表している。この事実をもう少し客観的に捉え、これを拡張することを考える。

表 2 は、それぞれの要因について、第 3 列が現状の GUI での表現、また第 4 列がその発展形となっている。ここで特筆すべきは現状の GUI の方で、われわれは、これを GUI の特徴として主張してきた。例えば音を表現するのに、波形やスペクトルで表現することは自明であり、ことさら声を大にして主張する必要がない。したがって、第 4 列のみに着目すると、なんら主張にならない。しかし、第 3 列の発展形として捉えると、新たな意味づけが行われる。以下、本 GUI の発展形をいくつか考察する。

同表で、1 項目だけ発展形とさせる音表現について考えてみよう。以下では、直接的に新たな音楽表現、あるいは音表現となる事例のほか、間接的に貢献しうるものについて述べる。間接的とは、音や音楽の整理法が提言でき、これにより、新たな音／音楽表現が誕生しやすい環境を提供する、というものである。

4.6. 対象：複合音、音楽

ひとつの音のみでなく、複合音を 1 つのオブジェクトで表すことはさまざまな場で行われている。MIR (Music information Retrieval; 音楽検索) の分野では、音楽そのものを 1 つのオブジェクトで表す方法などもある。後藤らは MusicCream で円板状の表示で 1 つの音楽を表している [後藤 03]。小坂は音合成システム Otkinshi (おっきんしゃい; 音と声と一緒にするシステム) の GUI において、音楽と音をつなぐため、1 つの楽

曲を1つのオブジェクトで表現し、この下にその音楽の持つ階層構造を表現した [小坂 01]。すなわち、譜面で書ける音楽、あるいは、サウンドをミックスして構成される音楽は階層的に記述できるため、これを表示されるオブジェクト上にも対応させ、オブジェクトをダブルクリックすると、その下位の階層のオブジェクトが表示される仕組みとなっている。すなわち、音楽もハイパーテキストと同様な記述とした。

このように、音や音楽を階層的に表し、これを粒子として描いたところで、何の新たな表現に繋がらないと思われるかもしれない。しかし、音楽のような上位のレベルに適用するより、音の下位レベルに適用させることにより新たな表現が行える。Schwarz の CBCS では、1つの音を小さな点で表現し、データベース上の音を2次元上の点で表現する。例えば、データベース上に1000個の音データがあると、なんらかの物理的特徴量を用いて、2次元上に1000個の点の散布図が描かれる。

この図上でマウスで範囲指定を行うと、その範囲内の音のミックスされた音がリアルタイムで計算されて出力される。マウスを動かすとそれに応じてリアルタイムに音が出る。これは、音データの音とは異なる音を得られるため、新たな表現たりうる。すなわち、数多くの音データの中から、どれを音響処理対象とするかどうかを瞬時瞬時にシステムに伝えることによる方法である。音響処理も、単純なミックス以外にも、補間など他の演算も考えられ、新たな音表現の可能性が考えられる。音楽にも適用した場合を考えると、新たなリミックス法の提案、といえる。

4.7. 表示法： 任意形状、波形、スペクトル

現表示法である球体という制約をはずすと、2次元としての円形、あるいはその他の形状、3次元としても球体以外の任意形状がある。また、波形やスペクトルなど、音の物理的実体を表現することもできる。しかし、別の発展法として、ひとつの音と具体的なオブジェクトを対応させない、という考え方もあろう。3項とも関連するが、たとえば、海の波面のように、3次元上の波の動きと音とを対応させることも考えられる。これは、光が粒子か波か、という問題と似ていて、粒子とする方がとっつきやすいが、波とする考えもある。

4.8. 時変／時不変の別：時変

電子音色辞書では、音のなんらかの物理的特徴量を3次元で表し、この布置上に音に対応する粒子を配置している。現在は音楽や音声認識で用いられる MFCC を用いている。この布置は時不変であるが、時変としてみよう。まず、1つは音そのものに動的要素があるので、

これを3次元上に表現することが考えられる。

例えば、振幅は瞬時に変化している。これを3次元上に表現すると、例えば、打弦楽器や撥弦楽器のように減衰系の音は、左から右に、あるいは奥行き方向に粒子が動く、という表現が可能になる。

4.9. 布置：任意に定義

音の物理的特徴量とは別に布置を定義することにより、音の関係をユーザの意図を反映させたものとして表現できる。例えば、音楽の変遷を年代順に並べた表記を行うことにより、音楽発生の相互関係を理解しやすい。これと同様に、自分にとっての整理しやすさの便宜で音や音楽の布置を決めることにより、音や音楽の統合のしやすさが飛躍的に向上する。

5.1 で述べた CBCS の場合もいえるが、数多くの音や音楽を瞬時に指し示すことは、システムがリアルタイムに複雑な音響信号処理が可能となり、質的な変化をもたらしうる。

以上は、4つの要因をただ1つだけ変更した例を述べたが、このほか、これらの複合要因を変化させることも考えられる。たとえば、3,4項を同時に発展させてみると、音の布置を任意として、さらに時変とする。

このような作曲法は、既に事例がある。古川作品 [古川 02] では、音のオブジェクトやこれらの演算オブジェクトが定義されており、これらが空間内を動くことにより、例えば衝突などの物理的インタラクションが生まれ、このイベントにより新たに音が生成され、音楽表へと発展する。これは音楽と視覚イベントとが対応するため、その意味でわかりやすい音楽となる利点がある。

5. サウンドスケープへの応用

舞台芸術としての音楽を離れ、環境と必ずしも切り離さない、音風景としてのサウンドスケープへの応用について考察する。4章で紹介した GUI は、さまざまな音情報の中継を行うための GUI に応用することができる。

小林らは、西表島や京都の瑞春院、東大本郷キャンパスの三四郎池などからの遠隔地環境音とリアルタイムに繋がる遠隔操作型 AUDIO I/O の開発を行った [Kobayashi 09]。このような遠隔操作型の音は音楽への応用が期待できるが、ここではこれには触れない。そのほかサウンドスケープへの応用が考えられる。4章で述べた Sound Planet は、これらの遠隔地から送られた音の中継点での GUI としての応用が可能である。

図4には、遠隔地から送られてきたさまざまな音情報が集まっているセンタ／公共の場／家庭での音の統合・選択・配信システムの GUI である。同図では、左側は

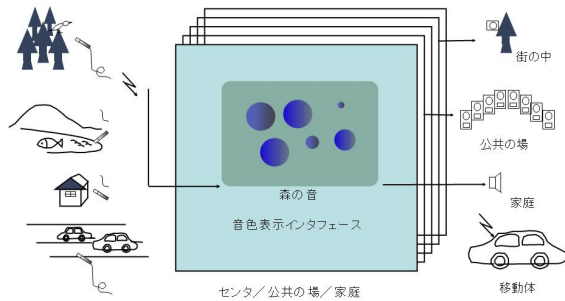


図 4. 音色表示インタフェースを用いた音情報の選択、処理、送信

遠隔地からの音入力を示し、右側が、選択、統合などの加工された結果の出力あるいは配信先を示したものである。

同図内の GUI では、1つのフレームが遠隔地の音源全体を表現し、また、その中の1つの粒子は、その中に含まれる個々の分離された音源を表すこともできる。もちろん、これは音源分離技術が適用されていることが前提である。

このような手法を用いて、音で環境を探るだけでなく、音により新たな環境をリアルタイムで提供するためのツールとして強力な機能を発揮することが期待できる。

6. まとめ

音データベースをネット上に登録、検索、合成する電子音色辞書をまず紹介した。これは音の Wikipedia ともいえるが、既存の freesound に比べて、音色記号を用いている点、またこれを用いた音色の逆引きを可能としている点が特徴である。またこのシステムで用いている、音表現法に3次元の GUI を紹介した。また、次にこの GUI の発展形から想起される新たな音あるいは音楽表現法について紹介した。個別の表現は決して新しいものではないが、本 GUI の発展形として整理して紹介した。特に、CBCS に見られるように、数多くの音を一度に指し占めす機能、あるいは動的なものを効率よく指し示す機能、は音響処理の入力情報として非常に重要で、この複雑なものを大量にまた、時変な対象物を簡単な操作で表現することにより、質的な変化をもたらすことができる、と期待している。今後は、ここでの構想の具体化を行っていく。

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。

7. 参考文献

- [Fernstom 02] Fernstom, M. and Cook, P.: Enhancing sonic browsing using audio information retrieval, in *Proc. of the ICAD 2002* (2002)
- [Hunt 80] Hunt, M., Lennig, M., and Mermelstein, P.: Experiments in syllable-based recognition of continuous speech, in *Proc. of the 1996 ICASSP* (1980)
- [Kobayashi 09] Kobayashi, H., Ueoka, R., and Michitaka, H.: Human Computer Biosphere Interaction: Toward a Sustainable Society, in *Proc. of the 27th international conference extended abstracts on Human factors in computing systems, SIGCHI* (2009)
- [mus] The freesound Project, Music Technology Group, <http://freesound.iua.upf.edu>
- [Osaka 09] Osaka, N., Saito, Y., Ishitsuka, S., and Yoshioka, Y.: An Electronic Timbre Dictionary and 3D Timbre Display, in *Proc. of the ICMC 2009* (2009)
- [Schwarz 08] Schwarz, D., Cadars, S., and Schnell, N.: What next? Continuation in Real-time corpus-based concatenative synthesis, in *Proc. of the ICMC 2008* (2008)
- [Tzanetakis 00] Tzanetakis, G. and Cook, P.: 3D Graphics tools for sound collections, in *Proc. of the DAFX-00* (2000)
- [吉岡 09] 吉岡 靖博, 石塚 慎也, 斎藤 佳紀, 高橋 奈緒子, 小坂 直敏: 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, 情報処理学会音楽情報科学研究会 MUS-79-11 (2009)
- [古川 02] 古川 聖: 「見ること, 聞くこと」-ASN (Active Symbol Notation) によるパフォーマンス ASN 制作: 古川聖, 石橋素 (コンセプトプログラミング), CG:: Robert Darroll, (財) 関西文化学術研究都市推進機構主催けいはんなメディアフェスティバル (2002)
- [後藤 03] 後藤 真孝: SmartMusic KIOSK: サビ出し機能付き音楽試聴機, 情報処理学会論文集, pp. 9-16 (2003)
- [小坂 01] 小坂 直敏, 榊原 健一, 引地孝文: Windows 上の音合成システム「おつきんしゃい」の構築, 通信学会誌 J84-D-II, No. 6, pp. 946-954 (2001)
- [小林 05] 小林 洋平, 小坂 直敏: 楽曲制作のための音色理論の構築にむけて, 音日本音響学会講演論文誌 (春), pp. 3-7-15 (2005)
- [小林 08] 小林 洋平, 小坂 直敏: 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), pp. 449-452 (2008)

著者プロフィールは、本会会報 Vol.1 No.1 を参照。