

出品作品について
— 音のモルフィング —

小坂 直敏

NTT基礎研究所
〒243-01 神奈川県厚木市森の里若宮 3-1
osaka@siva.brl.ntt.co.jp

あらまし モルフィングは映像処理の分野で一般的な技術となっているが、近年、音合成の分野でも研究が活発になってきた。本報告では、音のモルフィング技術の楽曲への応用例としてフルートとコンピュータのための音楽作品「鏡石」を題材として、そのモルフィング使用事例を述べた。ここでは、フルート音から歌声までとフルート音からプレート（打撃音）への二つのモルフィング音を中心とした例を述べた。またモルフィングの方法として、正弦波重畳モデルを基にフレーム単位でパラメータ補間を行う基本的な枠組を述べた。無声部で調波の存在しない区間を含む音も扱うことができ、単音の場合のモルフィング技術の拡張にもなっている。

キーワード モルフィング、コンピュータ音楽、ピッチシフト、音色補間、正弦波重畳モデル

On a submitted music piece

— Sound morphing —

Naotoshi Osaka

NTT Basic Research Laboratories
3-1 Morinosato Wakamiya, Atsugi-shi, Kanagawa 243-01 Japan
osaka@siva.brl.ntt.co.jp

Abstract Morphing has become a general technique in computer graphics. Recently the topic becomes popular as well in the field of sound synthesis. This paper describes application of the morphing technique to a music piece named "Mirror stone" for a flute and a computer. Two examples of morphing are introduced; from a flute to singing voice, and from flute to struck plates. Moreover, a basic framework of morphing technique based on a sinusoidal model is introduced, which framewise interpolates spectral parameters. The algorithm is an extended one of morphing in the case of mono-tone, since it can incorporate sounds with unvoiced parts, where there are no harmonics.

key words Morphing, computer music, pitch shift, timbre interpolation, sinusoidal model

1 はじめに

楽曲の創作の方法のひとつに、新しい音響技術を取り込み、これを全面に押し出して新しい音楽語法の主張とするものがある。

筆者らはこうしたコンピュータ音楽作成のため、また、マルチメディアコンテンツのための支援技術として自由な音色制御技術を検討している。これらの中で、現在特に二原音が与えられたときの音色補間問題(音色モルフィング)を扱っている。筆者はこれまで、これらの技術を応用してコンピュータ音楽作品を作ってきた。今回紹介するフルートとコンピュータのための楽曲では、音色モルフィングを音楽語法の中心に採り入れており、ここでは、この楽曲の構成、モルフィング技術の応用例などについて述べる。

2 楽曲の構成

ここで紹介する楽曲「鏡石」はフルートソロと NeXT コンピュータ (NeXT Turbo (2.8G Hard Disk)) と ISPW (IRCAM Signal Processing Workstation)[1] による音合成出力および実時間信号処理によるイフェクトとからなる。ブロックダイアを図1に示す。実演はフルート奏者と計算機操作者の2名で行う。

本曲は筆者の各種鏡を主題とした作品群の一環で、「真澄鏡」[P2]に継ぐ第2作で、1996年初演された[P3]。タイトルは R. Seymour ら作の物語 “Mirror Stone” 鏡石 (かがみいし)[2] にちなんでいる。

2.1 使用技術の概要

楽曲の構成を図2に示す。この図は全体を13のシーンに分け、各パートの役割を分類して表示したもので、同一の構成となるものをまとめて記したもので、ブロックダイアとスコアの折衷的な表示としている。この楽曲で用いられる技術は以下のように分類される。

1. NeXT+ISPW および MAX による音の実時間信号処理
2. “O⁴kinshi” (Oto to koega isshoni naru Shisutemu; おっきんしゃい) [3],[4] による音素材の生成、加工、編集、および演奏
3. モルフィングによる音色の補間技術

曲の大部分(シーン1,2,4,5,8,12)はフルートソロに対し、この拡声とイフェクトをかけるものである。リバーブ、イフェクトなど音の実時間信号処理は音楽用グラフィカルプログラミング環境 MAX[5]での機能を実現させた。ここでのイフェクトはピッチシフトによる多声化で、単なる3度シフトの他、ポリフォニーへのシフトなどがある。またこれらの音程を実時間で制御できるようにした。

フルート奏者の演奏に先立って奏者自身のフルート音を収録する。これをオフラインで加工し、合成音とした。これら音素材の生成、加工、編集、および演奏は O⁴kinshi によった。

同システムの音生成・編集部では収録された音(音声)の切り貼りの他、各種信号処理(フィルタリング、各種歪生成、サンプリング周波数変換)、合成音の作成を行い、サウンドファイルの形で予め蓄えておく。演奏部では表示されるボタンの押下により対応するサウンドファイルが発音する。なお、O⁴kinshi は、ミュージック コンクレートの思想に基づく楽曲創作支援用の汎用音楽システムを目指している。

以下では本楽曲で用いられているモルフィングとその技術を中心にして述べる。

3 モルフィングと楽曲への応用

モルフィングはCGでは多用されており一般的になっているが、音では現在まだ検討段階である。本来音色の連続的な変化自身は音楽表現では非常に一般的である。フォルテ、ピアノなどの表情は音量のみならず、音色も異なることが容易に理解される。そのため、クレッシェンド、デクレッシェンドなどの音量の変化に対する指定も、結果的には音色の連続的な変化を伴っている。

こうした音楽表現を拡張する意味で、演奏不可能な音色変化を音合成により実現する音のモルフィングは大きな期待が寄せられる。

特にある楽器から別の楽器へのモルフィングなどが適当な問題と考えられる。今回は、flute の音を中心にしてこれ分解したり、分解したものをまた組み合わせる元の音にする、他の楽器の音色への移行など、ここでは、flute の音色そのものの加工がテーマである。

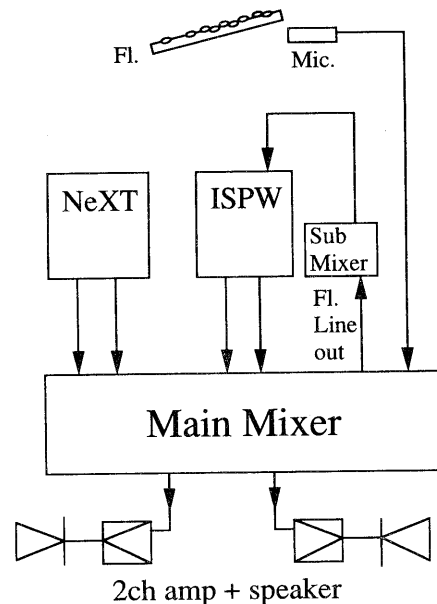


Fig.1 Block Diagram of "Mirror Stone" Performance System

なお、音のモルフィングは、高度な演奏技術として演奏者が行なうこともできる。図2のシーン6,10内フルート奏者によるMorph1はフルート音から地声、あるいはその逆を発音するもので、伝統的な管楽器奏法からは逸脱するが、なめらかな音色の移り変わりの実現のためには高度な演奏技術が要求される。

3.1 モルフィングの実例 1

図3はシーン6および10における譜例で、モルフィング音を中心として音楽が構成されている。フルート奏者は先に述べた演奏技術としてのモルフィング（譜例中 *m*）を行う。また計算機側はオフライン処理されたモルフィング音を提示する。計算機側はフルート音内でのなめらかなピッチシフト（グリッサンド）、歌声への（からの）モルフィング音を発音する。両者が交互に音を出し、重ね合わせながら音色が変化していく設計としている。譜中の音長にはあまり情報がなく、なめらかな音色の移行を主張しているのみである。テープ音楽（記録）とするときはこの部分全体がひとつのストリームに聞こえるよう意図しているが、実演奏では、スピーカ、聴者の位置などにも大きく左右され、ここのモルフィング音が重ね合わさっていく印象となる。

また、この部分はカデンツァのようなもので、モルフィング技術のデモとなることを想定しており、発表の回を増すごとにモルフィングの内容およびその品質が向上することを狙いとしている。

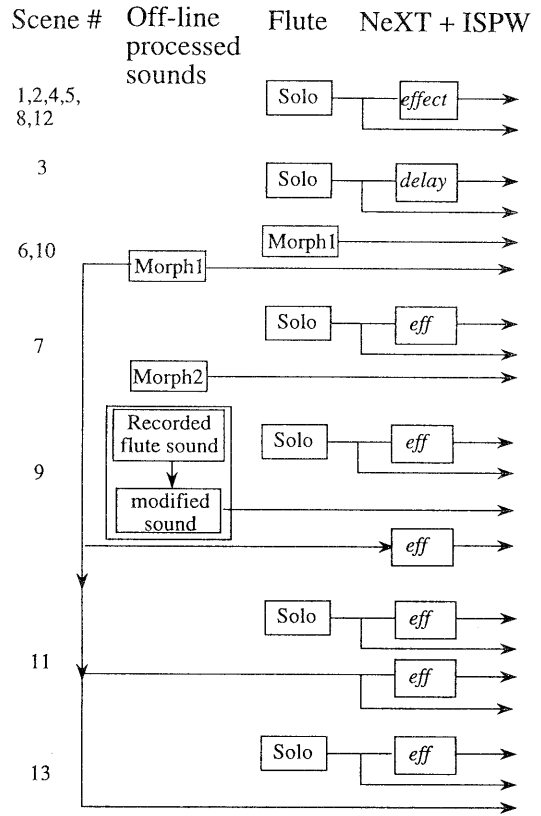


Fig. 2 The musical structure of "Mirror Stone"

Fig. 3 **Mirror Stone** Scene 6

The musical score for Scene 6 of "Mirror Stone" consists of three systems of staves. Each system has a Flute staff and a Comp. (Computer) staff. The notation includes various dynamics and performance instructions:

- System 1:** Flute and Comp. parts both feature morphing (*m*) markings.
- System 2:** Flute part starts with *m*, followed by a muraiiki (*) marking and *ff* dynamics. Comp. part starts with *m*, followed by *ff* dynamics and *m gliss* markings.
- System 3:** Flute part starts with *m*, followed by a muraiiki (*) marking. Comp. part starts with *m*, followed by *m*, a muraiiki (*) marking, *m*, and finally *gliss*.

Legend:
m: morphing *: muraiiki

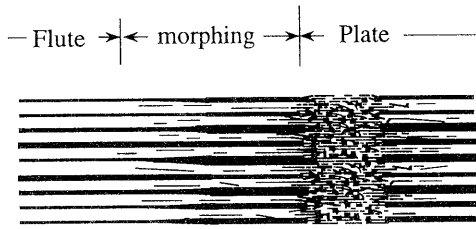


Fig. 4 Time sequence of flute to plate sound morphing

3.2 モルフィングの実例 2

図4はシーン7でのモルフィング例の概念図である。これは収録されたフルート音から単一ピッチが知覚できるプレートの打撃音へのモルフィングである。収録音は譜中のFlute部およびPlate部である。打撃音は冒頭に非定常部を含み、調波構造は明確に見えない。調波構造のある音から打撃音で始まるものへのモルフィングは、問題自身をどう設定するかとも関連する。ここでは、プレート音の打撃による非定常音が消滅し、調波構造が明確になった音を切出し、その部分へのモルフィングとした。

4 モルフィングの実現

モルフィングを実現するための方法は、Tellmanらによる楽音での音色補間の報告[6]など、近年いくつかの報告がある。これらの楽音の音色補間研究では、信号モデルとして正弦波重畳モデルが一般的に用いられてきた。これらの方式では音質の良いデモ音の提示、ソフトの配布などなされているが、アルゴリズム記述、手動と自動の区別などが明確でなく、他研究者のトレースが困難である。

筆者はこれまで、正弦波重畳モデルを用いて二つの定常音声の間あるいは、音楽的表現のない定常楽音を対象にした音色補間について報告してきた[7]。ここでは、さらに対象を拡張し、非定常音を含む楽音を対象に補間音色を求める方法について述べる。

4.1 基本的な考え方

音楽的表現を有す楽音はさまざまな音楽知覚的要素がある。単音でも知覚できる音色とは独立にピッチ、あるいはそれを基にメロディなどの要素がある。さらに、これらはピッチのない非定常部を含み、振幅あるいはピッチ変化を有すビブラート、強弱の変化などの音楽演奏上諸表情のある素材である。このような楽音の音色補間を以下の考えで行う。

1. 音色を連続的に制御できる。
2. 単音の音色補間は独自に満たす。
3. 二原音に共通する知覚要素は補間音も保存する。
4. 二原音に一方に存在し他方に存在しない要素は

対応する時点を算出し、存在しない原音の方を零レベルの要素として補間を行う。

1. は音色補間の定義である。2. はこれまでの方法と別な方法の提案であり、以下に述べる。3. は音色補間問題の基本的な設定である。二原音が共通なメロディ、音声では共通な音韻、文章であることなどである。補間音色も共通な要素を保ちつつ合成できるか否かが性能の重要な点である。なお、本技術の応用時には、オーケストラの合奏から街の騒音への音色補間のように1.だけの要素のみ関係するものも多い。

4.2 単音の音色補間

これまでの方法は、正弦波重畳モデルを用いて二原音それぞれを正弦波状のストリームとして知覚的に分解し、二つで対応するストリームを選び、そのパラメータを補間するものであった。

この方法で一定の性能が得られることは確認できた。しかし、処理がフレーム同期でなく、常に全時間を参照して対応関係を求めているため、原音が長くなればなるほど計算機のメモリが多く必要となり、演算上実際のでない。ここで提案する方法は演算の順序を変え、フレーム毎に合成パラメータを作成するもので、ストリームは合成時に結果として形成される。

5 音色補間モデル

5.1 スペクトルパラメータ表現

二原音の他、入力値として、二つの原音への近さを表す定数 $\alpha(0 \leq \alpha \leq 1)$ を定義する。また $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_{N_x})$ として時変として指定すると、音色が時間とともに変化する、いわゆるモルフィングの効果が実現できる。

これらの原音を正弦波重畳モデルを用いて表現し、原波形 x のスペクトルパラメータ \mathbf{X} について以下が出力される。

$$\mathbf{X} = \{\mathbf{X}_n \mid n = 0, 1, \dots, N_x - 1\} \quad (1)$$

$$\mathbf{X}_n = \{(f_{xl}^n, A_{xl}^n, \theta_{xl}^n) \mid l = 0, 1, \dots, L_x(n) - 1\} \quad (2)$$

ここに、 f は瞬時周波数で、瞬時位相 θ の時間微分である。また、 A は部分音 l の瞬時振幅を表す。添字 n はフレーム番号、また、 l は部分音番号である。 N_x は原音 x の総フレーム数を、 $L_x(n)$ はフレーム n の部分音総数表す。ここでは f, A, θ はそれぞれ境界値であり、

$$f_{xl}^n = f_{xl}^n(m = 0), \quad A_{xl}^n = A_{xl}^n(m = 0), \quad \theta_{xl}^n = \theta_{xl}^n(m = 0) \text{ と表される。}$$

同様に原波形 y のスペクトルパラメータについても以下で表される \mathbf{Y} が算出される。

$$\mathbf{Y} = \{\mathbf{Y}_n \mid n = 0, 1, \dots, N_y - 1\} \quad (3)$$

$$\mathbf{Y}_n = \{(f_{yl}^n, A_{yl}^n, \theta_{yl}^n) \mid l = 0, 1, \dots, L_y(n) - 1\} \quad (4)$$

(1)~(4)式のパラメータ推定および合成法は M & Q アルゴリズム [8], [2] を用いた。

同モデルの分析条件は、24kHz サンプリング、5msec 固定フレーム長、ハミング窓（窓長固定）による 4096 点短時間 FFT である。ただし、今回の合成は約 6kHz までの情報とした。

5.2 ピッチ抽出

ピッチ抽出はケブストラム法に拠った。短い無声区間は補間するなど、手作業による修正も可能なものとしてなめらかなピッチの曲線を推定しておく。

ピッチ抽出の目的は、1) 二原音に共通する要素がメロディなどピッチに関係する要素の場合に時点の対応を取るのに必要なため、また、2) 調波毎の周波数帯に分割し、対応関係はゾーン間でのみ操作するためである。2) は、ゾーンを越えた対応関係を作らない、また演算の簡略化が行える、などの利点がある。

5.3 時点対応の算出

次に DTW (Dynamic Time Warping) を用いて時間の非線形伸縮によるマッチングを行い知覚的に同一現象が生成している時点の対応をとる。楽音では、同一メロディーの対応時点、音声では、同一音韻を発生した場所などである。この結果求める K 個の時点の対応情報を $C(k)$ と表す。ただし、

$$C(k) = (C_x(k), C_y(k)), k = 0, 1, \dots, K-1 \quad (5)$$

ここで

$$C_x(k) \in \{i \mid i = 0, 1, \dots, N_x - 1\} \quad (6)$$

$$C_y(k) \in \{j \mid j = 0, 1, \dots, N_y - 1\} \quad (7)$$

これは原音間の k 番目の対応フレームである。

5.4 目的信号のスペクトルパラメータ

次に目的信号のスペクトル作成について述べる。図 5 は目的信号の時点と補間係数 α が与えられたとき、目的信号の合成に必要な二原音のそれぞれの時点のスペクトルを入力として、二原音間のパラメータの対応を抽出し、音声合成を行なう計算過程を示している。初期値計算部と単フレーム補間音色漸化計算部の 2 つに大別できる。両部共通に、スペクトル対応探索部とスペクトル補間部との組合せを単位としたユニットが用いられる。

初期値計算部では、最初に Z_0 を算出する。後は単フレーム補間音色漸化計算部で Z_{n+1} の計算を行ないこの二つを入力として単フレーム補間音色合成部が動作し、第 n フレームの合成音色波形 z_n を出力する。この繰り返し演算を z のフレーム数分行なう。なお、初回以降 Z_n の計算は不用である。

単フレーム補間音色計算部内の基準時点算出部では、時点の対応情報である (5) 式の $C(k)$ を入力し、実際の補間スペクトル Z_{n+1} を計算するのに必要となる

二原音のスペクトルの参照フレーム番号を k_{min} を算出する。これは、 Z_n の計算を行うのに最も適した二原音の参照フレームを決定するもので、時点対応するフレーム $C(k)$ の中から α に応じた Z_n に最も近い組み合わせとさせる。

$$k_{min} = \arg \min_k |(1-\alpha) \cdot C_x(k) + \alpha \cdot C_y(k) - n| \quad (8)$$

$$k_{min} \in \{k \mid k = 0, 1, \dots, K-1\} \quad (9)$$

ここでは対応をみるための二つの原音からの参照フレームは $X_{C_x(k_{min})}$ および $Y_{C_y(k_{min})}$ で与えられる。

6 スペクトルパラメータの対応と補間

ここでは、非定常信号、無声部を含む一般的な信号を対象とする。

6.1 ピッチとパワーの補間

まず、二原音の時点の対応点に沿ってピッチおよびパワーを補間し、目的信号のピッチとパワーを得る。これらは目的信号のスペクトルを得るための指針となるデータである。

6.2 スペクトルの対応探索

スペクトル対応探索部では要素数の異なる二つのグループ (異次元ベクトル) でその要素間に距離が定義できるとき、全体として最適に合致するよう相手を決定している。これは相手が見つからないことのコストを

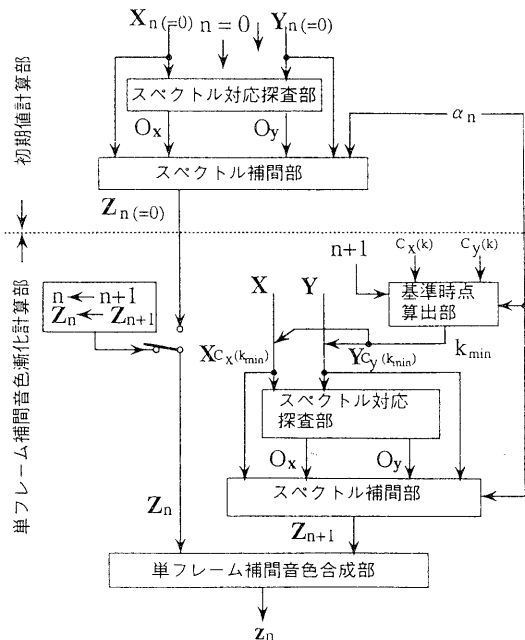


図 5 単フレームの補間音色計算方法

距離と同一次元で定義すると総合コスト最小の考えを用いて DP により決定でき、アルゴリズムの詳細は [7] に示した。

スペクトル対応を見るための距離尺度として、瞬時周波数を基本周波数で正規化し、調波番号に対応する数値 (実数) を用いた。また、この数値を指標として信号を調波と、残差の 2 種類に分け、クラス毎にパラメータの対応関係を調べた。

調波を特に分離するのは、調波番号による管理により、異音を生じさせないためである。また、スペクトル周波数を調波番号毎のゾーン (調波帯) に分割し、同一番号の調波帯間でのみ部分音の対応関係をみる。

こうして得られた対応関係を O_x, O_y で示す。これらも $C_x(k), C_y(k)$ と同様、対応する相手を示すポイントである。

なお、原音に無声部が含まれる場合は、以下の方法で調波帯に分割する。

1. 無声部と無声部: 便宜的に任意の調波帯幅を二原音共通に組む。

2. 無声部と有声部: 無声部の調波帯を有声部の調波帯と同一に設定する。

これらの調波帯分割はピッチ系列が無音も含むため時系列としては連続ではない。しかし、合成時にはフレーム n と $n+1$ は調波帯の概念をもたせず全帯域でパラメータの対応を見るため、この不連続の直接的な影響はない。

6.3 スペクトル包絡の補間

二原音の対応する時点のスペクトル包絡についても補間する。この補間は [9] などの考察もあるが、ここでは、簡易な手法としてピーク周波数とそのレベルをベクトルとし、これらの距離を用いて上記のアルゴリズムを適用した。対応が求まったものに対してはピーク周波数値とレベルを補間し、帯域のレベルがこの値となるよう各部分音の振幅を調整する。

スペクトル補間部では、対応がついたスペクトル同士は α の値により物理補間を行なう。また、対応がつかない場合はゼロ値への補間を行なう。こうして新しい音色のスペクトル Z_{n+1} が作成される。

6.4 合成部

単フレーム補間音色合成部では、このスペクトルと前のフレームで計算された Z_n との間で、再び [7] の方法で対応関係を見つけないことによりフレームと次フレームのスペクトルローカルピークの接続を行い、他は M&Q アルゴリズムによる瞬時位相関数を 3 次式で表現する合成法により、最終的に新しい音色波形 z_n を算出する。

7 あとがき

音のモルフィングを楽曲に応用した例として、96 年 7 月初演のフルートとコンピュータのための音楽「鏡石」を題材として、そのモルフィング使用事例を述べ

た。ここでは、フルート音から歌声までとフルート音からブレードの打撃音への二つのモルフィング音を中心とした譜例を述べた。またモルフィングの技術として、正弦波重畳モデルを基にフレーム単位でパラメータ補間を行う基本的な枠組を述べた。単音の場合では検討しなかった無声部を含み調波の存在しない区間への拡張にもなっている。今後はビブラートなどの音楽的表情が付加されたときの音のモルフィングについても検討し、これを応用する楽曲「鏡」シリーズをさらに充実させていく予定である。

謝辞 日頃討論していただく石井健一郎情報科学部長に感謝いたします。

参考文献

- [1] E. Lindemann et al., "The architecture of the IRCAM musical workstation," *Computer Music Journal*, pp. 41-50, Vol. 15, No. 3, Fall 1991.
- [2] Richard Seymour, Michael Palin and Alan Lee, "The Mirrorstone", Jonathan Cape Ltd, London, 1986. 邦訳 掛川恭子, 岩波書店, 1989.
- [3] 小坂直敏, 「楽音生成、演奏システムについて」, 信学技報 EA92-31, pp. 27-32, 1992. 6.
- [4] Naotoshi Osaka, "Otkinshi: A sound generation and performance system," *Proc. of ICMC 92*, pp. 406-407. San Jose, California, 1992.
- [5] M. Puckette, "Combining event and signal processing in the MAX graphical programming environment," *Computer Music Journal*, pp. 68-77, Vol 15, No. 3, Fall 1991.
- [6] E. Tellman, L. Haken and B. Holloway, "Timbre morphing using the lemur representation," *ICMC94 Proc.*, pp. 329-330, Sep. 1994.
- [7] 小坂直敏, 「Sinusoidal model による音色の補間」, 信学技報 SP95-130 pp.9-16, 平 8.2.
- [8] R. McAulay and T. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. on A.S.S.P.*, vol. ASSP-34, No. 4, Aug. 1986.
- [9] 白木善尚, 「時間的変化を考慮した音声スペクトル系列の補間とその評価」 音学講論, '94 春, 3-8-11, pp.251-252, 1994-3.

作品発表

- [P1] Naotoshi Osaka, "Prosody++," *Japan Today (MUSIANA'95)*, Louisiana Museum, Denmark, 1995.
- [P2] 小坂直敏, 「真澄鏡」, JACOM 主催, コンピュータ音楽の現在 '94, 神戸ジーベックホール 1994. 11
- [P3] 小坂直敏, 「鏡石」, JACOM 主催, コンピュータ音楽の現在 '96, 神戸ジーベックホール 1996. 7.