

擦弦楽器の意図表現合成のための奏法モデル

小泉 悠馬^{1,a)} 伊藤 克亘^{2,b)}

受付日 2012年7月2日, 採録日 2013年1月11日

概要: 擦弦楽器の合成音の柔軟な音色制御のための、物理モデルとスペクトルモデルのハイブリッドな分析合成系である、擦弦楽器の奏法モデルを提案する。また、モデルのパラメータを決定するための、演奏データから奏法による音色変化の特徴を分析する手法についても提案する。実測された演奏音のスペクトルから、奏法によって変化する調波擦弦振動、発音非調波擦弦振動、持続非調波擦弦振動の変化を分析しモデルパラメータを推定する。調波成分は、奏法によらないグローバルな擦弦振動を変化させ、知覚に関する乖離を補正する伝達特性として推定し、非調波成分は発想記号と基本周波数によって発生確率の変化する確率モデルとして奏法モデルを構築する。評価実験では、推定されたモデルに奏者の音楽的な表現意図が含まれるかどうかを擦弦楽器経験のある被験者5人が主観評価し、提案手法の有効性を評価した。提案法により推定されたモデルを用いることにより、実演奏音の発想記号による音色のイメージを保ったまま分析合成および音高制御をすることが可能であることを示した。

キーワード：擦弦楽器, 楽音合成, 楽音分析, 演奏表現, 音色制御

A Bowing Model for Synthesis of Expressive Intentions on Bowed String Instruments

YUMA KOIZUMI^{1,a)} KATUNOBU ITOU^{2,b)}

Received: July 2, 2012, Accepted: January 11, 2013

Abstract: This paper proposes a hybrid analysis-synthesis method that generates bowing models based on physical models and spectral models of bowed string instruments. To determine the model parameters, methods are also proposed for using performance data to analyze the characteristics of variations in tone color according to the bowing style. A bowing model consists of a harmonic component model, an onset inharmonic model, and a sustained-part inharmonic model. The harmonic component model is extracted by decomposing the actual performance spectra using the basic string motion spectra and resonance property of the instrument, and the inharmonic models consist of stochastic models. To evaluate the synthesized sound quality obtained by using the proposed model, five musicians compared actual sounds and synthetic musical sounds. No significant differences were observed between the actual sounds and the sounds generated by the proposed method.

Keywords: bowed string instruments, sound synthesis, sound analysis, performance expression, sound timbre control

1. はじめに

楽器の奏者は、その曲想や個性を活かすために、楽譜情報をもとに楽音をイメージし、それを奏法に変換する。特に意図表現は、発想記号と呼ばれる演奏の表現方法を指示する譜面上の記号をもとにイメージされ生成される。擦弦楽器の演奏音は発話のように明確な言語情報・意味情報を

¹ 法政大学大学院情報科学研究科
Graduate School of Computer and Information Sciences,
Hosei University, Koganei, Tokyo 184-8584, Japan

² 法政大学情報科学部
Faculty of Computer and Information Sciences, Hosei University,
Koganei, Tokyo 184-8584, Japan

a) 12t0005@cis.k.hosei.ac.jp

b) it@fw.ipsj.or.jp

持たないため、音量や音高の変化のほかに、音色の変化を用いて意図表現を伝達する。そのため、コンピュータによる表現力豊かで高品質な楽音合成においても、自由な音色制御が必要である。その実現のために、さかんに研究が行われている [1], [2], [3] が、人間のような演奏表現はまだまだ困難である。

擦弦楽器の代表的な楽音合成方式には、力学的センサを取り付けた楽器を用いて演奏することで、奏法情報を数値的に取得し、楽音を合成する物理モデル方式 [4] や、事前に用意した楽音コーパスから、合成したい音色に近いコーパスを取り出し意図表現を実現する素片接続方式 (e.g., *Vienna Symphonic Library**1) があげられる [5]。物理モデル方式は、奏者の意図表現による奏法情報を直接合成音に反映できるため、柔軟な合成が可能である。しかし、擦弦楽器演奏の熟練度が低い奏者は、意図表現に沿った擦弦振動の制御を行うことができないため、イメージする意図表現を忠実に合成することは難しい。一方、素片接続方式では、スペクトルモデル [6] により周波数領域で音色制御を行うため、自由に音色を変化させることが可能である。しかし Bonada ら [5] は素片接続方式を「コーパス楽音を演奏した奏者が生成しうる音響空間の生成」としており、意図表現は構築されたコーパスに依存する。また、擦弦楽器の演奏音には、奏者により制御された擦弦振動に、楽器の共鳴特性が畳み込まれており [7]、自由度が高すぎる周波数領域の操作では、奏者の意図表現と関係のない楽器の共鳴特性を操作してしまう可能性があるため、擦弦振動の成分のみを制御する合成手法が必要である。そのため、柔軟かつ意図表現による音色を制御するためには、物理モデルのようなその楽器の物理的な特性や奏者の意図表現の方法などをふまえたモデル化と、素片接続方式のような音色を容易に制御できるパラメータが必要となる。それを実現するためには、対象となる楽器の物理機構を考慮した演奏データの分析技術が必要となる。

本稿では、擦弦楽器の合成音の柔軟な音色制御のための、物理モデルとスペクトルモデルのハイブリッドな分析合成系である、擦弦楽器の奏法モデルを提案する。本稿では、奏者の意図表現情報は擦弦振動に表れ、聴衆はその情報を演奏音のスペクトルの変化という形で知覚しているという仮定の下で、グローバルな擦弦振動と実演奏音との乖離を埋める形で奏法モデルが導出される。奏法モデルでは、奏法による調波構造の変化、発音時の非調波成分の変化、持続区間における非調波成分を制御する。モデル評価のために、複数の発想記号に対して合成音を生成し、提案手法の有効性を主観評価によって評価する。

以下では、2章で擦弦楽器の楽音生成過程について述べる。次いで、3章で提案する奏法モデルを導入・構築法を

説明し、4章で提案法を用いた楽音の合成手順を説明する。最後に実音声と提案法により合成された合成音を比較することにより、提案法の有効性を示す。

2. 擦弦楽器音の生成過程

擦弦楽器では弦の振動が駒を通して楽器本体で共鳴することにより楽音が生成される [8] (図 1)。奏者は、意図表現のために、弦を抑える左手でビブラートをかけ、弦を擦る右手で様々な奏法を駆使し弦の振動を制御する。奏法によって擦弦の位置・圧力・速度の制御方法が変化し、これらのパラメータの相対的な強度差により、弦の各モードの振動の強さや、非周期成分の強さなどが決定され、特徴的な音色が生成できる。

本章では、実演奏音から擦弦振動の成分の変化を解析するために、擦弦楽器の物理現象について考える。

2.1 擦弦振動

擦弦中の基本的な弦の運動は、Helmholtzにより、駒と枕により固定される弦が描く放物線上をなぞる三角波として知られている [9]。この振動はヘルムホルツ振動と呼ばれ、Stick-Slip 運動と呼ばれる、弓に弾かれた弦が臨界点まで引っ張られ、臨界点に到達すると滑り、また摩擦により弓に引っ張られるという現象により生成される調波成分にのみパワーを持つ振動である。ヘルムホルツ振動 $h(t)$ の変位の式を次に示す。

$$h(\beta, t) = \sum_{n=1}^N A(\beta, n, t) \sin(\omega_n t + \theta_n) \quad (1)$$

$$A(\beta, n, t) = b(t) \frac{\sin(n\pi\beta)}{n^2} \quad (2)$$

ただし、 $\omega_n = 2\pi n F_0(t)$ ($F_0(t)$ は基本周波数)、 θ_n は n 番目のモードの位相 $\theta_n = n\pi$ 、 N はナイキスト周波数までに存在しうる振動のモード数、 β は擦弦位置 x と弦長 l の比率 x/l 、 $A(\beta, n, t)$ はヘルムホルツ振動における各モードのパワー、 $b(t)$ は時刻 t における振幅である。

しかし、実際に観測される弦の振動は、奏法による各モードの振動の強度の比率の変化 [10] や、弦のスティフネスの効果などによる不規則振動 [11] が含まれおり、三角波ではない。この不規則振動には、擦弦運動の非線形性や、非周

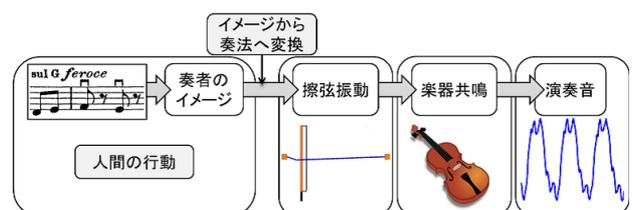


図 1 擦弦楽器の音の生成過程

Fig. 1 The process of sound generation for bowed-string instruments.

*1 <http://vsl.co.at>

期成分の間欠性や slip 現象の周期からのランダムなずれと関係するカオス理論 [12] が関係しているとされており、未解決の部分も存在する。

また、発音時の弓圧と加速度が釣り合わない場合、数十 ms 程の不安定な slip 現象が発生し、振動ごとの周期が安定しない擦弦振動が発生する [13]。この不規則な slip 現象には、1 周期中に複数回の slip が起きる multiple-flyback や、本来 slip が起きる位置よりも遅れて slip が起きる prolonged periods が存在する。multiple-flyback では、“loose/slipping sound” と呼ばれるひっくりかえったような音が生成され、また、prolonged periods では“choked/creaky sound” と呼ばれるカリカリとしたノイズが発生するといわれている。特に後者は、marcato（はっきりと）や feroce（荒々しく）などの、音に迫力を付与する意図表現の演奏の際に用いられる [14]。

2.2 楽器の共鳴

擦弦楽器音は、擦弦振動が駒を通して楽器へ伝達され、楽器内で共鳴し、音色変化することによって生成される。物理モデルにおいて、楽器の共鳴による擦弦振動の変化は、楽器を線形時不変系と仮定され、インパルス応答により計測される [15], [16]。

3. 奏法モデルの構築

擦弦楽器の演奏において、人間が直接制御を行う部分は擦弦振動である。したがって、意図表現による音色変化は擦弦振動に表れると仮定でき、物理モデルのように擦弦振動における奏法の特徴を制御することで、音色を柔軟に制御できる。しかしながら、物理モデルでの楽音制御には実際の奏法情報が必要であり、力学的センサを用いないとコントロールが難しい。

そこで本章では、生成楽音の音色を制御するために、周波数領域で音色を制御する奏法モデルの導入を提案し、またその構築法について説明する。擦弦楽器の演奏音からの意図表現の知覚には、各モードの振動強度の変化による調波成分の変化と、発音部の非調波成分が深く関係する [14], [17]。奏法モデルは、実演奏音から奏法による擦弦振動のスペクトルパラメータの変動を推定することにより構築され、式 (1) からの意図表現の知覚に関するスペクトルの乖離を埋める。

擦弦振動のスペクトルには、奏法の影響について線形で表現可能な成分と、非線形でないと表現できない成分が存在する。調波成分の変化は、奏法による式 (1) からの各モードの振動強度の変化ととらえることができるため、線形な伝達系として記述できると考えられる。しかし、意図表現に深く関係する発音ノイズやスティフネスの効果による非調波成分は、スペクトルの概形が非定常かつ、周波数成分の分布もランダムであり、カオス理論を考慮する必要

があるため、線形な伝達系としては記述できない。

そこで、調波モデルはヘルムホルツ振動により生成される三角波からの変化を表す線形な伝達系として記述し、非調波成分は式 (1) とは別個に考え、非調波発音モデルは発想記号と基本周波数、持続部非調波モデルは発想記号によって非調波成分の発生確率の変化する確率モデルとして奏法モデルを構築する。

3.1 調波モデル

観測された演奏データは、擦弦振動に楽器の共鳴特性が畳み込まれている。そこで、演奏データに楽器の共鳴特性の逆フィルタをかけ、擦弦振動のみを取り出す前処理を行う。ここで観測される擦弦振動は、式 (1) の各モードの振動が奏法により変化を起し、さらに非調波成分が加算されたものと見なすことができる。よって式 (1) は以下の式のように書き換えることができる。

$$s(t) = \sum_{n=1}^N A(\beta, n, t) E(n, t) \sin(\omega_n t) + \psi(t) \quad (3)$$

ここで $E(n, t)$ は調波モデルであり、奏法による第 n モードの振動変化である。また、 $\psi(t)$ は非調波成分を表す。

さらに、式 (3) を短時間フーリエ変換 (STFT) して求めるスペクトログラムは以下ようになる。

$$S(\omega, k) = \sum_{n=1}^N A(\omega_n, k) E(\omega_n, k) + \Psi(\omega, k) \quad (4)$$

ここで ω 周波数、 k は時刻に対応するインデックスを表す。

$S(\omega, k)$ をピッチ同期の分析手法である TANDEM [18] により求めることにより、持続擦弦区間などの $\Psi(\omega, k)$ が十分に小さい区間においては、調波成分 $A(\omega_n, k) E(\omega_n, k)$ の真値が求まる。そこで、周波数軸方向に連続なスペクトログラム $S(\omega, k)$ から、調波成分の値のみをサンプリングすることにより、調波成分にのみ値を持つスペクトログラム

$$S(\omega_n, k) = A(\omega_n, k) E(\omega_n, k) \quad (5)$$

を求める。ここで TANDEM 窓に用いる基本周波数 $F_0(t)$ は、河原らの手法 [19] を用いて求めた。

本研究で取り扱う調波モデルは、音色制御を取り扱うものであるため、 $E(\omega_n, k)$ は信号のパワーを変化させないものとする。また、式 (1) の基本ヘルムホルツ振動は $1/\beta$ のモードを持たない。そこで、倍音強度が十分に小さくなる第 30 倍音までのパワーを基本ヘルムホルツ振動に持たせるために、 $\beta = 1/31$ に固定する。すると、時刻 k における基本ヘルムホルツ振動 $A(\omega_n, k)$ のパワーは以下の式で求まる。

$$\sum_N |A(\omega_n, k)|^2 = \sum_N |S(\omega_n, k)|^2 \quad (6)$$

ここで、基本ヘルムホルツ振動の各モードの振幅は、式 (2)

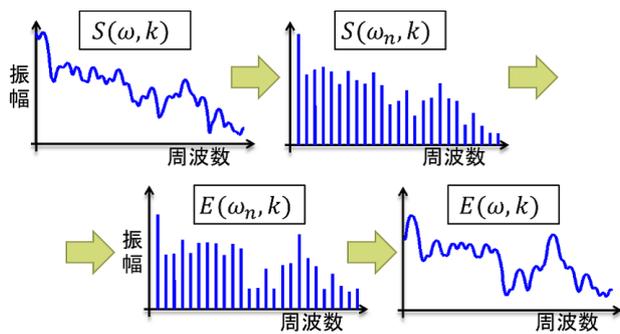


図 2 調波モデルの構築

Fig. 2 Constructing the harmonic component model.

で定義されるため、式 (6) は以下のように書き換えることができる。

$$\sum_N \left| b(k) \frac{\sin(n\pi\beta)}{n^2} \right|^2 = \sum_N |S(\omega_n, k)|^2 \quad (7)$$

よって、インデックス k における基本ヘルムホルツ振動の振幅 $b(k)$ は、以下のように求めることができる。

$$b(k) = \sqrt{\frac{\sum_N |S(\omega_n, k)|^2}{N \sum_N \left| \frac{\sin(n\pi\beta)}{n^2} \right|^2}} \quad (8)$$

TANDEM 窓作成の際に推定された $F_0(k)$ と、固定された β 、および式 (8) より、ヘルムホルツ振動のスペクトログラム $A(\omega_n, k)$ を求めることができる。よって奏法による調波成分への影響である調波モデル $E(\omega_n, k)$ は以下の式で求めることができる。

$$E(\omega_n, k) = \frac{S(\omega_n, k)}{A(\omega_n, k)} \quad (9)$$

$E(\omega_n, k)$ は調波成分にのみ値を持つ、各フレームのスペクトルに対する係数行列である。そこで、窓関数を用いて $E(\omega_n, k)$ を補間し、さらに基本周波数の幅で振動する成分を平滑化により取り除くことにより [20]、周波数軸方向に連続で滑らかな調波モデル $E(\omega, k)$ を推定する (図 2)。

3.2 非調波発音モデル

Guettler ら [14] は、非調波発音区間を、雑音ではなく楽音と知覚する限界を 50 ms としており、多くの楽音においてその長さが 50 ms 以下であることを示した。そこで本稿では発音区間を、音の立ち上がりから 45 ms と定義する。

発音時の非調波成分は、押弦位置で定義される基本周期 \hat{F}_0 と異なる周期で発生する slip 現象に起因する。そのため、適切な長さでない切り出し窓を用いて STFT を行うと、発音区間のスペクトルはピンクノイズのようなスペクトルとして観測される。これは、毎周期ごとに slip の位置が異なり、1 波長分の長さが変わるため、分析区間に様々な周波数成分が混在するためである。しかし発音区間であっても波形は Stick-Slip 現象により生成されているた

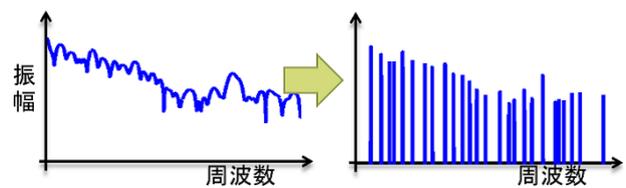


図 3 発音時のスパースなスペクトルの推定

Fig. 3 Onset sparse noise spectrum estimation.

め、1 波長分のスペクトルは周波数軸方向にスパースであると考えられる。

よって発音時に生成される、短い区間ではスパースなスペクトルの生成モデルを構築することで、発音時の不安定な slip 現象を合成することができる。そのため、演奏データから発音時に発生している周波数成分を推定しモデル化する必要がある。

しかし、毎波長ごとに周期が変わる波形に対し、ピッチ同期の STFT を行うことは難しく、また、STFT は切り出し窓長が短い場合周波数分解能が悪い。そこで波形の切り出しには、押弦位置で決定する基本周波数の 4 倍のハンニング窓を用いて STFT を行う。解析されたスペクトルからピーク検出を行い、ピークの値のみをサンプリングしてスペクトルをスパース化する (図 3)。

発音時のスペクトログラムを生成するにあたり必要なパラメータは、発生する周波数成分の数 N_f とその周波数 ω_f である。本稿では、発生する周波数成分の数 N_f を平均 μ 、分散 σ^2 の正規分布に従う確率変数、周波数 ω_f を $\mathcal{U}(\omega_f \hat{F}_0/2, f_s/2)$ の一様分布に従う確率変数と見なす (f_s はサンプリング周波数)。

ここで観測データから構築すべきモデルパラメータは μ と σ^2 である。発音区間の 1 波長分の長さの変化は押弦位置により決まる波長の長さより長いものと短いものの 2 種類と考えられるため、発生する周波数成分 N_f の数は、平均 $2N$ の分布に従うと仮定する。また、発生する周波数成分の数が N となると、それは調波擦弦振動であると考えられるため、 $N_f > N$ であると考えられる。よって、正規分布の性質から、標準偏差は $\sigma < \frac{N}{6}$ であると考えられる。

本稿では、 μ の事前分布は、平均 $2N$ 、分散 $\sigma^2/3$ の正規分布、 σ^2 の事前分布は、平均 $(N/12)^2$ 、分散 1 の逆ガンマ分布を仮定し、観測データからベイズ更新を用いて事後分布を求めた。

3.3 持続部非調波モデル

持続部の擦弦ノイズは特に未解決部分の多い問題であるが、slip 現象時に弦のステイフネスの効果により発生する雑音であることが分かっている。物理モデル [21] では物理パラメータにより生成された擦弦振動 $\tilde{s}(t)$ に対し、以下のような確率モデルでノイズが付与された擦弦振動 $s(t)$ 生成

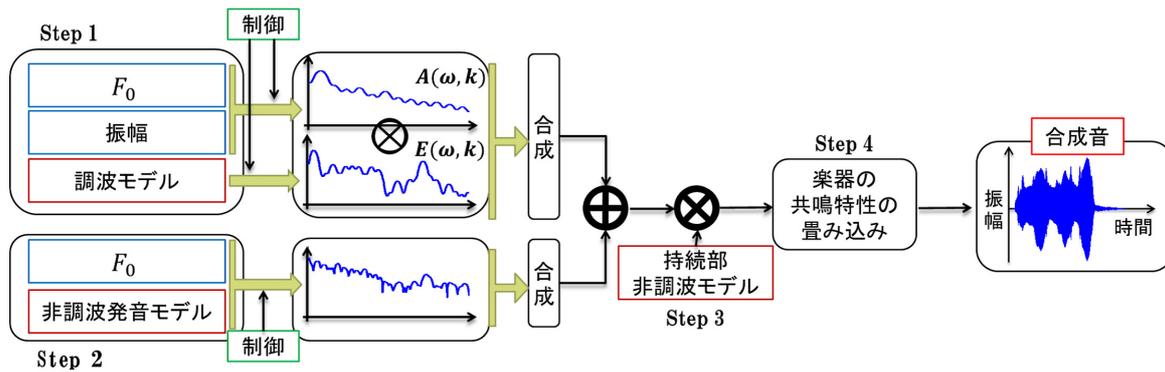


図 4 奏法モデルを用いた楽音合成
Fig. 4 Sound synthesis using bowing models.

している。

$$s(t) = \begin{cases} (O + G(t)u(t))\tilde{s}(t) & (\text{slip}) \\ \tilde{s}(t) & (\text{stick}) \end{cases} \quad (10)$$

ここで $u(t)$ は 0 から 1 の範囲の一様分布の乱数, O はノイズ強度のための任意の定数, $G(t)$ は時間変化するノイズの強度である。本稿では, $G(t)$ を持続部非調波モデルとして扱う。

式 (4) より観測された擦弦振動スペクトルにおける非調波成分 $\Psi(\omega, k)$ は, 観測擦弦振動のスペクトログラム $S(\omega, k)$ から, 調波振動スペクトル $A(\omega_n, k)E(\omega_n, k)$ を周波数軸方向に窓関数で補間したスペクトルを減算することで求めることができると仮定する。よって持続部非調波モデル $G(t)$ は

$$G(k) = \sqrt{\frac{1}{2\tau_k} \sum_{\omega} |\Psi(k)|^2} \quad (11)$$

で求めた $G(k)$ を時間波形長に補間することで求める。ここで τ_k は k 番目のフレームの STFT 時の切り出しフレーム長である。

4. 奏法モデルを用いた合成音の実験

構築した生成モデルの有効性の評価のために, 奏法モデルを用いた分析合成音と制御合成音の品質を, 主観評価によって評価する。本章ではまず, 奏法モデルを用いた楽音合成法について述べ, さらに実際の擦弦楽器演奏者を対象とした意図表現と品質の聴取実験を行う。

4.1 楽音の合成

奏法モデルを用いて楽音を合成する (図 4)。本稿では, モデルを分析した演奏音を分析合成, または制御する手法について議論する。

Step1: 調波擦弦振動の生成

まず, 基本擦弦振動のスペクトログラム $A(\omega, k)$ を作成する。 $A(\omega, k)$ は, 倍音比率が固定のスペクトログラムな

ので, 3.1 節で推定された基本周波数 $F_0(k)$ と振幅 $b(k)$ から生成することができる。ここで, $F_0(k)$ と $b(k)$ を制御することにより, 音高の操作や楽音全体の音量を制御する。

次に生成された $A(\omega, k)$ に対し, 調波モデル $E(\omega, k)$ を掛け合わせ, 任意の調波擦弦振動のスペクトログラム $S(\omega_n, k)$ を生成する。これは式 (3) における調波モデルのフィルタリングと見なすことができる。

最後に $S(\omega_n, k)$ に Overlap-Add 法を用いて, 調波擦弦振動 $\tilde{s}_{\text{harm}}(t)$ を合成する。

Step2: 非調波発音擦弦振動の生成

非調波発音擦弦振動が知覚において重要となる発想記号 (e.g., marcato, feroce) は非調波発音モデルを用いて発音擦弦振動を合成する。

発音時の不安定な slip 現象により発生する周波数成分 ω_f は, 押弦位置で決定する基本周波数 \hat{F}_0 の倍音成分の前後 $F_0(t)/2\text{Hz}$ に $N_h = N_f/N$ 個発生する周波数成分である。よっての各調波成分周辺に発生する周波数成分の範囲は $(n\hat{F}_0 - \hat{F}_0/2 \leq \omega_f < n\hat{F}_0 + \hat{F}_0/2)$ となる。しかし, N_h は周波数の個数なので自然数をとる必要があるが, 正規分布である非調波発音モデルより生成される周波数成分の数 N_f は実数値であり, $N_h = N_f/N$ では N_h も実数値をとる。よって各調波成分周辺に発生する周波数成分の数 N_h は, 以下の式によって自然数へと変換される。

$$N_h = \alpha + \gamma \quad (12)$$

ここで, $\alpha = \text{floor}(N_f/N)$, $\beta = N_f \bmod N$, γ は平均 β のベルヌーイ分布に従う確率変数である。

生成されたスパースなスペクトログラムを窓関数を用いて周波数方向に補間したものを, Overlap-Add 法を用いて合成する。さらにここで合成された発音擦弦振動と, Step1 で合成された調波擦弦振動 $\tilde{s}_{\text{harm}}(t)$ を加算合成することにより, 擦弦振動 $\tilde{s}(t)$ を生成する。

Step3: 持続擦弦振動雑音の付与

生成された擦弦振動に, 持続部非調波モデルと式 (10) に

基づきノイズを付与し、 $s(t)$ を生成する。slip 区間において、擦弦振動の微分値は負の値となる。そこで、生成された擦弦振動 $\tilde{s}(t)$ に対し微分を行い、slip 区間を判定する。

Step4: 楽器の共鳴特性のフィルタリング

生成された $s(t)$ に対し、楽器の共鳴特性をフィルタリングする。本稿では、先行研究 [1] で計測された共鳴特性と同様のものを用いた。

4.2 評価実験

構築された奏法モデルの有効性の検証のために、合成音の主観評価実験を行う。実験では、発想記号による自身の意図表現を、楽器を用いて忠実に再現することが可能な熟練度を持つ奏者が、3種類の発想記号 (feroce: 荒々しく, marcato: はっきりと, dolce: やわらかく) で演奏したバイオリンの4種類の音高の単音 (G線のG音: 197 Hz, G線のB音: 234 Hz, A線のA音: 442 Hz, A線のC音: 525 Hz) を用いた。すべての演奏音は、ICレコーダを用いて、防音室で録音した。収録条件は、標準化周波数は 48 kHz, 量子化 bit 数は 24 bit とした。発想記号は、物理モデルにおける奏法パラメータの変化が特徴的かつ, marcato と dolce はオーケストラや室内楽の楽曲に頻出するものであり, feroce は音色の迫力やアクセントが特徴的であり, dolce と極端な音色の差を持つ, という理由で選択した。

本実験では、実演奏音 (ORG), 提案法 (PRO), 提案法を用いて音高を短3度上昇させた合成音 (INT+3) と下降させた合成音 (INT-3), 合成 Step1 において調波モデルを掛け合わせず, $S(\omega_n, k)$ に $A(\omega, k)$ を用いた合成音 (-EMF) と, 発音時の非調波成分が近くに大きく影響を及ぼすとされる発想記号である feroce と marcato においては, 合成 Step2 を省略し, 非調波発音モデルを用いなかった合成音 (-ATT) を追加した計 68 種類の楽音を用いた。dolce に関しては, PRO において, 非調波発音モデルを合成に用いないため, -ATT を評価しない。

被験者は、擦弦楽器を5年以上経験し、発想記号による音色の変化をイメージできる5名とした。音圧は、被験者の聴きやすいレベルとなるよう、事前に調節した。詳細な実験条件を表 1 に示す。

評価は、提示音の、自身の発想記号のイメージに対する音色の合致度および音色の自然さ (A) と、音質 (B) を、それぞれ5段階で評価する MOS (Mean Opinion Score) を用いて行った。各評定は、1が非常に悪い、2が悪い、3が普通、4が良い、5が非常に良い、を表す。刺激の提示順序はランダムとし、被験者にはどの刺激が合成音であるかは伝えずに評価した。各刺激の間には 3sec の間が空けられる。

4.3 評価実験の結果

MOS により算出された、各合成法の平均値と標準誤差

表 1 実験条件

Table 1 Experimental conditions.

発想記号	feroce, marcato, dolce
音高	197 Hz, 234 Hz, 442 Hz, 525 Hz
音長	feroce, dolce: 二分音符 (BPM = 120) marcato: 四分音符 (BPM = 120)
演奏者	バイオリン歴 12 年の大学生
収録機材	TASCAM DR-07 内蔵マイク使用
収録条件	48 kHz, 24 bit
収録部屋	空調を切った防音室
比較音	実演奏音, 提案法, 音高操作音 (短3度上昇, 下降), 調波モデルなし, 非調波発音モデルなし
被験者数	5 名
スピーカー	BOSE Companion 2 series II

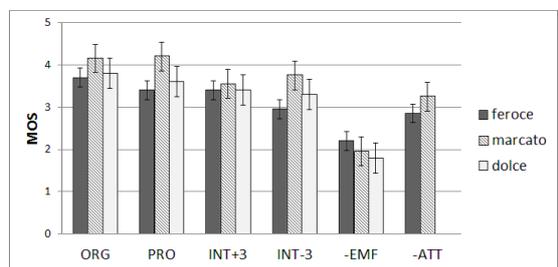


図 5 主観評価の結果 (イメージに対する音色の合致度および音色の自然さ)

Fig. 5 Result of the subjective evaluation (Naturalness of timbre and coincidence level between musical images and timbres).

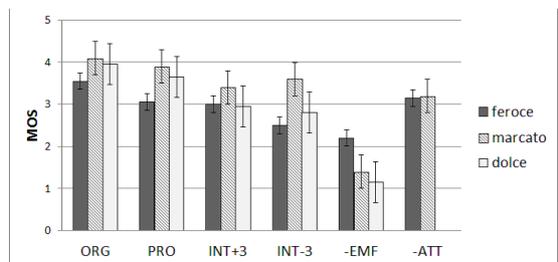


図 6 主観評価の結果 (音質)

Fig. 6 Result of the subjective evaluation (Sound quality).

を、発想記号ごとに図 5 と図 6 に示す。図の横軸は合成法の種類、縦軸はイメージに対する合致度および音色の自然さ (図 5) と音質 (図 6) を示す。

図 5 から、提案法の評点は実演奏音と比べ若干の低下が見られるが、ほぼ等価であることが見て取れる。Dunnnett の多重比較検定 [22] により有意差を検定した結果、提案法による合成音は、すべての発想記号において実演奏音と比較して危険率 5% で有意差が認められなかった。また音高操作を行った合成音においても、すべての発想記号において実演奏音と比較して危険率 5% で有意差が認められなかった。調波モデルを用いなかった合成音は、すべての発想記号において実演奏音と比較して危険率 5% で有意差

が認められ、発音非調波モデルを用いなかった合成音は、*marcato* において実演奏音と比較して危険率 5%で有意差が認められた。

この結果から奏法モデルを用いた楽音合成は、実演奏音の発想記号による音色のイメージを保ったまま分析合成および音高制御をすることが可能であると分かる。また、発音非調波モデルを用いない分析合成では、*feroce* においては有意差は認められなかったが、*feroce*, *marcato* ともに音高操作を行った楽音よりも評点が下がっていることから、発音非調波モデルが発想記号による音色のイメージの分析合成の品質に効果を持つと考えられる。また、調波モデルを用いない合成音に有意差が認められたことから、意図表現の知覚において音色の変化が重要な要素であることが示された。

音質 (B) の評価結果を、Dunnett の多重比較検定により有意差を検定した結果、提案法による合成音は、すべての発想記号において実演奏音と比較して危険率 5%で有意差が認められなかった。また音高操作を行った合成音においては、*feroce* の短 3 度上昇音と *marcato* において実演奏音と比較して危険率 5%で有意差が認められなかった。調波モデルを用いなかった合成音は、すべての発想記号において実演奏音と比較して危険率 5%で有意差が認められ、発音非調波モデルを用いなかった合成音は、*marcato* において実演奏音と比較して危険率 5%で有意差が認められた。

この結果から提案法を用いた楽音合成は、音高操作を行わない分析合成において実演奏音の音質を保ったまま分析合成することが可能であると分かる。しかし、音高操作を行う場合は、いくつかの発想記号において音質の劣化が生じる。(A) において有意差が認められず、(B) において有意差が認められることから、これは楽音合成の際、音高操作によって生じる部分的な位相の不整合に起因する問題であると考えられるため、今後合成法の改良が必要であると考えられる。

本評価実験では、単音のみを用いて評価を行ったが、提案法は奏法による音色変化を演奏音と同等の音質が得られる分析合成が可能のため、フレーズであっても同様の性能が達成されると考えられる。今後は、フレーズ内での発音非調波成分の分析法の検討を進める必要がある。

また本評価実験は、バイオリンの楽音のみを対象とした。提案法は、実演奏音を擦弦振動と楽器の共鳴特性に分解し、擦弦振動から奏法によるスペクトルパラメータの変動を推定する手法であるため、バイオリンと同じような物理機構を持つ、ビオラやチェロなどにおいても同等の性能が達成されると考えられる。また発想記号に関しては、提案法は、聴衆の耳になじみ深い *marcato*, *dolce* などや、特徴的な音色を持つ *feroce* などの発想記号に関して有効な合成手法であると考えられる。

5. おわりに

本稿では、擦弦楽器の合成音の柔軟な音色制御のための、物理モデルとスペクトルモデルのハイブリッドな分析合成系である、擦弦楽器の奏法モデルを提案した。奏法モデルを用いた分析合成音の発想記号のイメージに対する音色の合致度および音色の自然さと音質の評価では、バイオリンの単音では元音声に対し有意差のない楽音を分析合成できることを示した。また、構築した奏法モデルを固定した音高制御では、発想記号のイメージに対する音色の合致度および音色の自然さに関しては、元音声と同等の品質で楽音制御を行えることを示した。

しかしながら、本研究で扱った評価実験は、バイオリンの楽音のみを対象とし、発想記号は *feroce*, *marcato*, *dolce* の 3 種類のみを扱ったものであった。今後は、他の擦弦楽器や、網羅的に実験することは困難であるとしても他の多くの種類の発想記号においても分析合成を行い、評価実験を行う。

また提案法を、楽音合成システムとして用いることを想定した場合、奏法モデルの制御パラメータの検証も必要となる。今後、ユーザの所望の音色を合成するために、ユーザが直観的に扱いやすいパラメータを用いた制御手法を検討していく。

謝辞 本研究に対し、バイオリンの共鳴特性をご提供くださった Esteban Maestre 氏 (Center for Computer Research in Music and Acoustics) に感謝いたします。また本研究の一部は、科研費基盤研究 (C) の支援を受けた。

参考文献

- [1] Maestre, E., Blaauw, M., Bonada, J., Guaus, E. and Perez, A.: Statistical Modeling of Bowing Control applied to Violin Sound Synthesis, *IEEE Trans. Audio, Speech, and Language Processing*, Vol.18, No.4 (2010).
- [2] Serra, X.: Musical Sound Modeling with Sinusoids plus Noise, *Musical Signal Processing*, Leman, M. and Berg, P. (Eds.), Swets & Zeitlinger Publishers (1997).
- [3] Dannenberg, R.B. and Derenyi, I.: Combining Instrument and Performance Models for High-Quality Music Synthesis, *J. New Music Research*, Vol.27, No.3, pp.211–238 (1998).
- [4] Demoucron, M.: On the control of virtual violins: Physical modelling and control of bowed string instruments, Ph.D. thesis, Universite Pierre et Marie Curie (UPMC), Paris & Royal Institute of Technology (KTH), Stockholm (2008).
- [5] Bonada, J. and Serra, X.: Synthesis of the Singing Voice by Performance Sampling and Spectral Models, *IEEE Signal Processing Magazine*, Vol.24, No.2, pp.67–79 (2007).
- [6] Bonada, J. and Loscos, A.: Sample-based singing voice synthesizer by spectral concatenation, *Proc. Stockholm Music Acoustics Conference (SMAC)* (Aug. 2003).
- [7] Cremer, L.: *Physics of the Violin*, pp.201–382, The MIT Press, Cambridge, MA (1984).

- [8] McIntyre, M.E. and Woodhouse, J.: The acoustics of stringed musical instruments, *Interdisciplinary Science Reviews*, Vol.3, No.2, pp.157–173 (1978).
- [9] 近藤正夫：バイオリンの発音機構, 日本音響学会誌, Vol.49, No.3, pp.184–192 (1993).
- [10] 村上智之：擦弦振動の平均法による解析, 日本機械学会論文集, C編, Vol.62, No.598, pp.2102–2109 (1996).
- [11] McIntyre, M.E., Schumacher, R.T. and Woodhouse, J.: Aperiodicity in bowed-string motion, *Acustica*, Vol.49, pp.13–32 (1981).
- [12] Popp, K. and Stelter, P.: Stick-Slip Vibrations and Chaos, *Philosophical Transactions: Physical Sciences and Engineering*, Vol.332, No.1624, pp.89–105 (1990).
- [13] Guettler, K.: On the Creation of the Helmholtz Motion in Bowed Strings, *Acta Acustica united with Acustica*, Vol.88, No.6, pp.970–985(16) (Nov./Dec. 2002).
- [14] Guettler, K. and Askenfelt, A.: Acceptance limits for the duration of pre-Helmholtz transients in bowed string attacks, *Journal of the Acoustical Society of America*, Vol.101 (1997).
- [15] Cook, P.R. and Trueman, D.: A database of measured musical instrument body radiation impulse responses, and computer applications for exploring and utilizing the measured filter functions, *Proc. 1998 Int. Symp. Musical Acoust.*, Leavenworth, WA (1998).
- [16] Perez, C.A., Bonada, J., Patynen, J. and Valimaki, V.: Method for measuring violin sound radiation based on bowed glissandi and its application to sound synthesis, *J. Acoust. Soc. Am.*, Vol.130, No.2, 1020-9 (Aug. 2011).
- [17] Caclin, A., McAdams, S., Smith, B.K. and Winsberg, S.: Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones, *J. Acoust. Soc. Am.*, Vol.118, No.1, pp.471–482 (July 2005).
- [18] 森勢将雅, 高橋 徹, 河原英紀, 入野俊夫：窓関数による分析時刻の影響を受けにくい周期信号のパワースペクトル推定法, 電子情報通信学会論文誌 D, 情報・システム, Vol.J90-D, No.12, pp.3265–3267 (2007).
- [19] 河原英紀, 森勢将雅, 西村竜一, 入野俊夫：基本波の FM と AM 成分に基づく高速な基本周波数推定法について, 日本音響学会聴覚研究会資料, Vol.41, No.9, pp.679–684 (Dec. 2011).
- [20] Kawahara, H. and Morise, M.: Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework, *Sadhana*, Vol.36, Part 5, pp.713–727 (Oct. 2011).
- [21] Chafe, C.: Pulsed Noise in Self-Sustained Oscillations of Musical Instruments, *Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Albuquerque (1990).
- [22] Dunnett, C.W.: New Tables for Multiple Comparisons with a Control, *Biometrics*, Vol.20, No.3, pp.482–491 (1964).



小泉 悠馬 (学生会員)

2012年法政大学情報科学部デジタルメディア学科卒業, 現在, 法政大学大学院情報科学研究科情報科学専攻修士課程に在学中. 音楽音響信号の演奏分析合成の研究に従事. 2012年文部科学省第1回サイエンス・インカレ奨励表彰, 2012年日本音響学会第6回学生優秀発表賞を受賞. 日本音響学会学生会員.



伊藤 克亘 (正会員)

会員.

博士(工学). 1993年電子技術総合研究所入所. 2003年名古屋大学大学院情報科学研究科助教授. 2006年法政大学情報科学部教授, 現在に至る. 音声・音楽を主とした音処理と自然言語処理全般に興味を持つ. 日本音響学会