

物理モデルによるモーフィングにおける音色知覚

引地 孝文 小坂 直敏

NTT基礎研究所

〒243-0198 神奈川県厚木市森の里若宮 3-1

0462-40-3514 0462-40-3655

hikichi@idea.brl.ntt.co.jp osaka@idea.brl.ntt.co.jp

あらまし 本研究では、コンピュータ音楽およびマルチメディアコンテンツ創作支援技術として、音色の自由な制御方式を提案することを目的とする。この一環として、これまで物理モデルによる音色モーフィングを検討してきた。ここでモーフィングとは、二つの異なる音色の間を知覚的に「連続的に」「補間する」よう音合成することを指す。前回、一次元の弦を打撃、撥きにより励起する音合成モデルを用い、音色モーフィングの一手法を提案した。今回、この手法により合成された補間音を用い、音色類似性に関する評価試験を行った。その結果、比較的近い二音色について、パラメータの線形補間によりモーフィングの二条件を満たす結果が得られた。また、多次元尺度構成法による布置の図より求められるモーフィングについての評価尺度を提案した。

キーワード 音色制御、物理モデル、音色モーフィング、打弦、撥弦、多次元尺度構成法

On the Timbre Perception of Sound Morphing using a Physical Model

Takafumi Hikichi Naotoshi Osaka

NTT Basic Research Laboratories

3-1, Morinosato Wakamiya, Atsugi, Kanagawa 243-0198 Japan

+81-462-40-3514 +81-462-40-3655

hikichi@idea.brl.ntt.co.jp osaka@idea.brl.ntt.co.jp

Abstract Our goal is to implement a flexible sound timbre control system that can be used to compose computer music and multimedia content. For this purpose, we have investigated sound morphing based on physical modeling. Morphing is defined as synthesizing sounds which “interpolate” “smoothly” between two different timbres. In previous work, a method to morph sounds that used a sound-synthesis model composed of a one-dimensional string and the excitation of striking and plucking has been proposed. In this report, experiments are performed to evaluate the perceptual similarities among several tones synthesized by this method. The result for two relatively close tones shows that two morphing conditions, namely, “interpolativeness” and “smoothness” can be fulfilled by simple parameter interpolations. Further, two evaluation criteria calculated by the outcome of multidimensional scaling techniques are proposed.

key words timbre control, physical models, sound morphing, struck strings, plucked strings, multidimensional scaling

1 はじめに

筆者らは、コンピュータ音楽およびマルチメディアコンテンツ創作における音合成技術に貢献するため、音色の柔軟な制御を目的とした研究を行なっている[1][2]。これまでこの一環として、物理モデルによる音色モーフィングを検討してきた[3]。ここで音色モーフィングとは、二つの音色の知覚空間上における中間音色を、「連続性」および「補間性」の二条件を満たすよう合成することを指す。今回、打弦および撥弦を対象とした音色モーフィングによる合成音の知覚について、主観評価試験を行った結果を報告する。

前報告[3]では、励起体(exciter)により励起された弦、棒等の1次元の振動体(vibrator)の振舞いを表す力学モデルを用い、物理パラメータ操作によるモーフィングの一方式を提案した。物理モデルを利用する長所として、

1. 少ない物理パラメータで補間できる制御の容易さ
2. 中間の状況も物理的実体をモデル化できているため、両端と同等な音質の合成が可能

という点が挙げられる。本モデルのパラメータを後の実験に用いる値とともに表1に示す。これらのパラメータの補間により、音色空間上で連続性、補間性を満たすことが期待される。今回、モーフィングの二条件について検証するため、以下の実験を行った。

2 モーフィングの評価尺度

これまで補間音色に関する評価尺度の検討は少ない。そこで、モーフィングの性能を客観的に評価することを目的とし、モーフィングの二条件である「連続性」「補間性」に対する評価尺度を提案する。多次元尺度構成法により一連の刺激音に対する空間的な布置が得られた時、次のように定義する(図1参照)。ここで、刺激音の布置を以下の様に表す。

$$X_i = (x_{i1}, \dots, x_{ir}), i = 0, 1, \dots, N-1$$

(ただし N は刺激個数、 r は次元)

まず、布置におけるモーフィング両端の刺激音の距離 D は次式で表される。

$$D = \left[\sum_{a=1}^r (x_{0,a} - x_{N-1,a})^2 \right]^{\frac{1}{2}}$$

1. 連続性について

モーフィングが連続的であるとは、隣り合う刺激音の布置間の距離が等間隔であることとする。そこで、隣り合う刺激音の距離 $d_{i,i+1}$ の標準偏差を両端の刺激音の距離で正規化した以下の指標を考

える。

$$M_1 = \frac{\frac{1}{N-2} \sum_{i=1}^{N-1} d_{i,i+1}^2 - \left(\frac{1}{N-2} \sum_{i=1}^{N-1} d_{i,i+1} \right)^2}{D}$$

この M_1 が小さいほど各距離の偏りが小さく連続的であると考えられる。

2. 補間性について

補間性を満たすとは、補間音が両端の中間に布置することである。そこで、中間からのずれを測る指標として、両端を結んだ線分と各補間音の布置 X_i との距離 \hat{d}_i の平均値を D で正規化したものを考える。

$$M_2 = \frac{\frac{1}{N-2} \sum_{i=1}^{N-2} \hat{d}_i}{D}$$

M_2 が小さいほど補間性がよいと考えられる。

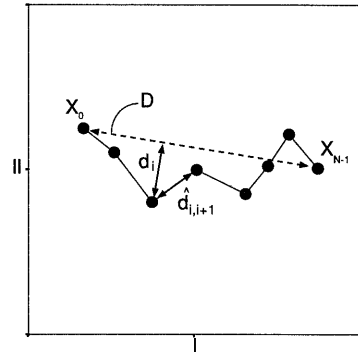


図1: モーフィングの評価尺度に関する説明図

3 実験1: 周波数包絡及び非調波性の影響

3.1 実験条件

ここではまず、1つのパラメータのみを操作した時の知覚空間を調べる。可変パラメータとして減衰係数 b_3 およびヤング率 E を用いる。その理由は、これらは音色の印象に大きく関係すると考えられるためである。 b_3 は周波数減衰包絡を、 E は非調波性を変化させる。

次に用いた刺激音について述べる。スチールギター弦の撥弦音を模擬した合成音に対し、減衰係数 b_3 を11段階に変化させた一連の刺激音を作成した。これを

表 1: 音合成モデルのパラメータと使用した値 (t_{nr} は励起体が自然に弦から離れるまでの時間長であり、他のパラメータより求められる。)

パラメータ	ナイロン弦 (NS)	ピアノ弦 (PI)	スチール弦 (SS)
励起体 (ハンマー)			
アタック相対位置 h	0.12	0.12	0.12
ハンマー質量 M_H [kg]	2.97×10^{-3}	2.97×10^{-3}	2.97×10^{-3}
指数係数 p	3.5	2.5	3.5
スチフネス係数 K	4.5×10^9	4.5×10^9	4.5×10^9
初速度 V_H [m/s]	10.0	10.0	10.0
力の打ち切り長 t_f [s]	$t_{nr}/2$	t_{nr}	$t_{nr}/2$
弦			
長さ L [m]	0.6135	0.62	0.6135
半径 a [m]	3.9×10^{-4}	5.0×10^{-4}	2.52×10^{-4}
密度 ρ [kg/m ³]	1.12×10^3	8.07×10^3	7.86×10^3
張力 T [N]	55.2	667	162
ヤング率 E [N/m ²]	2.5×10^7	2.0×10^{11}	2.0×10^{11}
減衰係数 b_1	0.5	0.5	0.8
減衰係数 b_3	3.5×10^{-8}	6.25×10^{-9}	3.0×10^{-9}
出力相対位置 o	0.7	0.7	0.7
離散化条件			
分割数 N	91	69	80
サンプリング周波数 f_e [kHz]	48	48	48

カテゴリ A とする。また、ピアノ弦の打撃音、スチールギター弦の撥弦音の模擬合成音に対し、ヤング率 E を 11 段階に変化させた刺激音を作成した。これらをそれぞれカテゴリ B、カテゴリ C とする。ヤング率 E は線形に、減衰係数 b_3 については対数を取った値を線形に変化させた。刺激音は 2 秒長である。表 2 に用いた刺激音と変化させたパラメータ値を示す。ここでは補間率 α ($0.0 < \alpha < 1.0$) を用いて刺激音を表す。

音色知覚に関する従来研究において、音色を対比較する際にピッチ、ラウドネスを厳密に揃える方法、これらも音色の定義に含め、揃えずに実験を行うなど様々な方法が採られている [4]。ここでは、刺激音のピッチ、ラウドネスに関して次の物理的制約を課す。すなわち、理論的な基本周波数が同一となるように合成パラメータを設定し、かつ信号パワーが等しい刺激音を比較する。

実験には一対比較法を用いた。それぞれのカテゴリ毎にランダムに選ばれた一対の刺激音を、ヘッドホン (STAX SR-A Pro.) を用いて継時的に両耳呈示し、被験者に音色の類似度を「同じ」(評価値 0) から「全く異なる」(評価値 6) の 7 段階で評価させた。呈示のタイムパターンを図 2 に示す。全ての刺激音についてカテゴリ A は 4 回、カテゴリ B, C は 6 回の繰り返し判断を行い、評価値の合計を求め非類似度データとした。被験者 1 人あたりの総判断回数は、カ

テゴリ A に対し 220 回 ($= {}_{11}C_2 \times 4$)、カテゴリ B および C のそれぞれに対し 330 回 ($= {}_{11}C_2 \times 6$) である。被験者は聴力の正常な成人女性 4 名である。

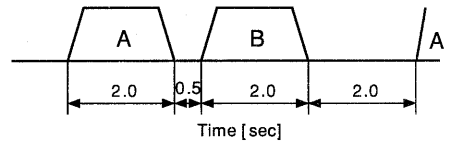


図 2: 合成音呈示のタイムパターン

3.2 実験結果及び考察

評価実験の結果得られた非類似度データを主観的距離と見なし、多次元尺度構成法 (Kruskal の方法) を適用し知覚空間上の配置を求めた。ストレスと次元数の関係を図 3 に示す。ストレスの肘は明確ではないが、2 次元においてもストレスの値は十分小さいためここでは 2 次元解を採用した。得られた布置を図 4~6 に示す。

図 4 において、各点は I 軸方向に拡がって分布し、I 軸が減衰係数の変化に対応している。 $\alpha=0.0 \sim 1.0$ を結んだ経路は II 軸負方向に湾曲し、また多少振動し

表 2: 実験 1 に用いた刺激音の補間率 (刺激番号) と変化させたパラメータの値

カテゴリ A		カテゴリ B, カテゴリ C
α	減衰係数 b_3	ヤング率 E
0.0	1.0×10^{-9}	0.0
0.1	1.58×10^{-9}	4.0×10^{11}
0.2	2.51×10^{-9}	8.0×10^{11}
0.3	3.98×10^{-9}	1.2×10^{12}
0.4	6.31×10^{-9}	1.6×10^{12}
0.5	1.00×10^{-8}	2.0×10^{12}
0.6	1.58×10^{-8}	2.4×10^{12}
0.7	2.51×10^{-8}	2.8×10^{12}
0.8	3.98×10^{-8}	3.2×10^{12}
0.9	6.31×10^{-8}	3.8×10^{12}
1.0	1.0×10^{-7}	4.0×10^{12}

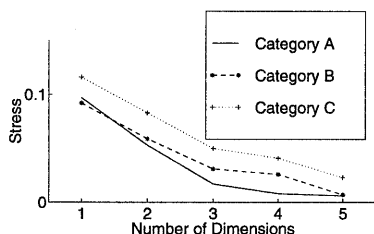


図 3: ストレスと次元数の関係

ているが (0.4, 0.5 付近)、ほぼ単調に推移し、概して滑らかに並んでいると考えられる。両端を結ぶ線分から最も離れている配置 0.4 までの両端を結ぶ線分との正規化距離は 0.4 である。用いた合成音の FFT スペクトルの一例を図 8 に示す。周波数成分の相対関係を保持しつつ、周波数減衰包絡の傾きが変化していることが分かる。

図 5 について、II 軸の正方向にふくらみ、かつ $\alpha = 0.0$ が他から離れている。これは、 $\alpha = 0.0$ 以外の刺激音では非調波性を持つことによるうなりが生じ、この知覚が $\alpha = 0.0$ とは異なるためであると考えられる。

一方図 6 においては、0.0 から 0.4 まではほぼ等間隔に推移するが、0.5~0.9 が縮退している様子が観察される。これは両端の音色の差が小さいことに起因すると推察される。(0.0, 0.1, 0.2, 0.4, 0.5, 0.8, 1.0) の系列を取ると、知覚的にほぼ等間隔に変化すると考え

られる。

表 7 に、音色知覚空間上の布置より求めたモーフィングの評価値を示す。カテゴリ B における補間性尺度 M_2 が突出しており、布置による補間経路の湾曲の度合いと直感的に一致する。

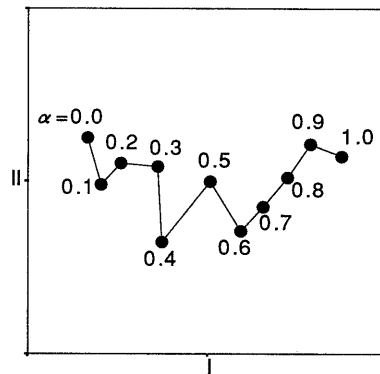


図 4: カテゴリ A (スチールギター撥弦) の刺激音に対する音色知覚空間 (ストレス: 5.3%, 2 次元解)

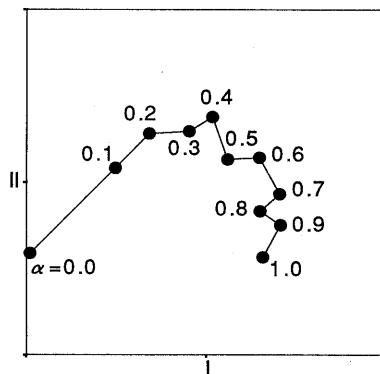


図 5: カテゴリ B (ピアノ打弦) の刺激音に対する音色知覚空間 (ストレス: 5.9%, 2 次元解)

4 実験 2: 補間音色の評価

4.1 実験条件

異なる物理パラメータをもつ媒質を想定し、この間の補間音色を合成することを試みる。まず、複数のパラメータを同時に補間する方法を以下に示す。これは

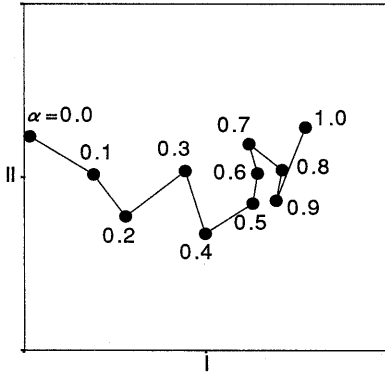


図 6: カテゴリ C (スチールギター撥弦) の刺激音に対する音色知覚空間 (ストレス: 8.3%, 2 次元解)

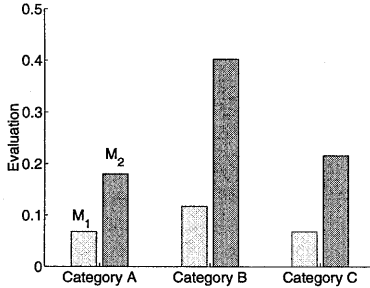


図 7: 音色空間の布置より求めたモーフィング評価値

弦の補間 2 を除き前報告 [3] と同一である (以下添字 X, Y は両端のパラメータを表す)。

弦 (振動体) の補間

1. 長さ、半径、密度、減衰係数をそれぞれ線形に補間

$$(L(\alpha), a(\alpha), \dots) = (\alpha L_X + (1 - \alpha)L_Y, \alpha a_X + (1 - \alpha)a_Y, \dots)$$

2. 両端での非調波性を表す係数 B を求め、これを線形に補間するようヤング率の値を決定

$$B_i = \frac{\pi^3 a_i^4 E_i}{8 L_i^2 T_i}, \quad (i = X, Y)$$

$$B(\alpha) = \alpha B_X + (1 - \alpha)B_Y.$$

$$E(\alpha) = \frac{8L(\alpha)^2 T(\alpha) B(\alpha)}{\pi^3 a(\alpha)^4}.$$

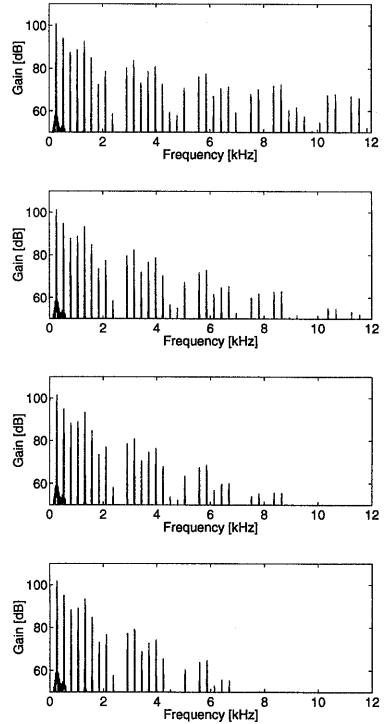


図 8: カテゴリ A の合成音の FFT スペクトル (上から、それぞれ $\alpha = 0.0, 0.1, 0.2, 0.3$)

3. 基本周波数を一定とするよう張力を決定

$$T(\alpha) = \{2L(\alpha)a(\alpha)f_1\}^2 \rho(\alpha)\pi.$$

4. 弦の離散化分割数 N を安定性の条件より導出

打弦と撥弦 (励起条件) の補間

励起のパラメータのうち、 M_H, K は固定とし、指数係数 p と力の継続時間 t_f を用いて補間する。ここでは $p = 3.5$ 、力の時間長 t_f は自然にリリースの生じる時間長の $1/2$ を撥弦での値とした。そして打弦、撥弦の間を p, t_f に関して線形に補間した。ここで、補間率 α が 1 の時に打弦、0 の時撥弦である。また、添字 st, pl はそれぞれ、打弦、撥弦のパラメータを表し、 t_{nr} は自然にリリースが生じる力の持続長を表す。

$$p(\alpha) = \alpha p_{st} + (1 - \alpha)p_{pl},$$

$$t_f(\alpha) = \alpha t_{nr} + (1 - \alpha)\frac{t_{nr}}{2} = \frac{1 + \alpha}{2} t_{nr}$$

このようにパラメータを操作した場合の外力波形の例を図 9 に示す。

4.2 実験結果及び考察

実験により得られた非類似度データを主観的距離と見なし、多次元尺度構成法 (Kruskalの方法) を適用し知覚空間上の配置を求めた。ストレスと次元数の関係を図10に示す。2次元においてストレスの値は十分小さいため2次元解を採用した。得られた布置(2次元解)を図11,12に示す。

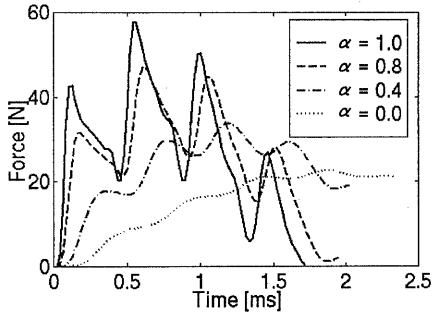


図9: パラメータ補間により補間された弦に加わる力の波形 (カテゴリ D)

上記のように求めたパラメータを用いて一連の刺激音 ($\alpha=0.0\sim 1.0$ の11段階) を作成し、実験1と同様に対比較試験を行った。端点としてスチールギター弦、ナイロンギター弦の撥弦と、ピアノ弦の打弦を仮定した。これらのパラメータ値は前掲の表1に示したものをを用いた (ただし SS はスチール撥弦、PI はピアノ打弦、NS はナイロン撥弦をそれぞれ表す)。刺激音のカテゴリ D、E を以下の様に定めた。

- カテゴリ D: 両端でのパラメータとしてスチール弦の撥弦 (SS) とピアノ弦の打弦 (PI) を使用
- カテゴリ E: 両端でのパラメータとしてナイロン弦の撥弦 (NS) とピアノ弦の打弦 (PI) を使用

被験者1人あたりの総判断回数は612回 ($= {}_{18}C_2 \times 4$)、その他の実験条件は実験1と同様である。

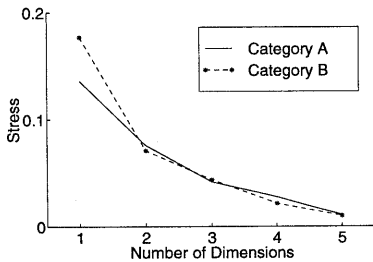


図10: ストレスと次元数の関係

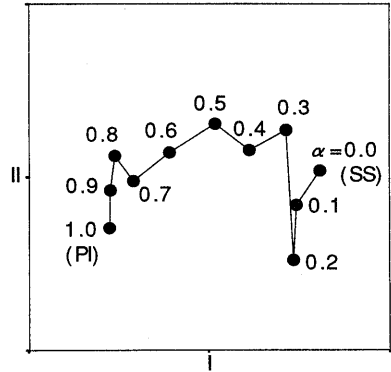


図11: カテゴリ D (スチールギター撥弦 SS とピアノ打弦 PI の補間) の刺激音に対する音色知覚空間 (ストレス 7.6%, 2次元解)

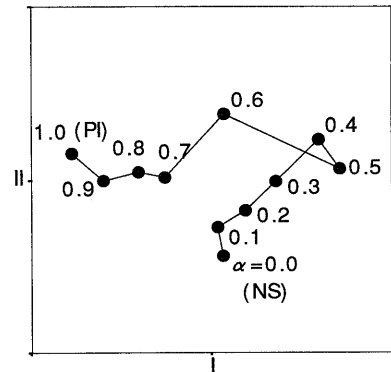


図12: カテゴリ E (ナイロンギター撥弦 NS とピアノ打弦 PI の補間) の刺激音に対する音色知覚空間 (ストレス 7.1%, 2次元解)

図11より、0.1, 0.2がI軸の負方向に飛び出しているが、これらを除けば連続性、補間性ともに比較的よく満たす。0.7~0.9は固まって布置する様子が見られる。

図12の布置は、 $\alpha=0.0$ と 1.0 を結ぶ線分方向と直

交する方向へ拡がっている。カテゴリ E の合成音の冒頭の 100ms を切出し、FFT 分析を行った結果、補間率 $\alpha=0.5$ 付近において、基本周波数の相対レベルが突出する傾向が見られた。また、周波数包絡の変化に規則性は見られなかった。図 14 にこの一例を示す。これは、単純な線形補間によって信号スペクトルは補間されないことを示している。

図 13 に、音色知覚空間上の布置より求めたモーフィングの評価値を示す。カテゴリ E の評価値は連続性 M_1 、補間性 M_2 ともに大であり、布置より得られる直感と合っている。実験 1 のカテゴリ B と比較すると、補間性の指標 M_2 はやや小さいが連続性の指標 M_1 はより大である。すなわち、補間性についてはわずかに勝っているが連続性に劣る。

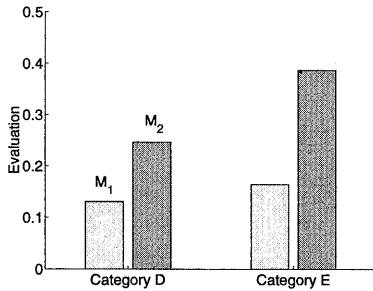


図 13: 音色空間の布置より求めたモーフィング評価値

5 まとめ

我々は、コンピュータ音楽、マルチメディアにおける音楽情報処理に資するため、音色の柔軟な制御方法を研究している。今回、これまでに提案した物理モデルによる補間音色合成法を用い、聴取実験による補間音の知覚特性について検討した。その結果、変化するパラメータが減衰係数、ヤング率単独の場合、物理モデルによるモーフィング音はほぼ連続的に、かつ概ね両端を補間することを示した。また、スチール弦とピアノに関しては良好な結果を得たが、ナイロン弦とピアノに関しては期待した結果が得られず、パラメータの補間方法について課題を残した。

今回、知覚空間上の経路が多少振動し、また布置が縮退をする傾向が見られたが、これは両端の知覚的距離が離れたものになれば緩和されると期待される。

モーフィングの評価に関して、客観的尺度を考案しその傾向をみた。結果は多次元尺度法による布置と直感的に合うものであった。今後事例を増やしさらに検討する予定である。

今後は、さらにパラメータ範囲を拡大した場合の検討、モーフィングに関する評価尺度の検討、および、異なるモデル間のモーフィング方法についての検討等が課題である。

付録

本稿で用いた音合成モデルは次式の微分方程式に基づいている。

$$\frac{\partial^2 y}{\partial t^2} = \frac{T}{\mu} \frac{\partial^2 y}{\partial x^2} - \frac{\kappa^2 ES}{\mu} \frac{\partial^4 y}{\partial x^4} - 2b_1 \frac{\partial y}{\partial t} + 2b_3 \frac{\partial^3 y}{\partial t^3} + f(x, x_0, t).$$

ここで、 y : 弦の変位、 μ : 線密度、 T : 張力、 E : ヤング率 (剛性を表す)、 κ : ジャイレーション半径、 S : 弦断面積、 $f(x, x_0, t)$: 外力密度、 x_0 : 打弦位置である。第三、第四項は減衰項であり、この定式化により次式で与えられる周波数に依存した減衰率 d を表現している。

$$d(\omega) = b_1 + b_3 \omega^2$$

ω : 角周波数である。

ここで用いる励起モデルは、パラメータ操作により打弦と撥弦の両方を表現でき、従来のモデルの拡張となっている。まず、撥弦体がある速度で弦と衝突し、弦と接触を保ちつつ平行位置から変位を生じる。接触の際には、打弦モデルと同様の反発力を生じるとする。ある時間経過の後、撥弦体は、弦からの反発力が 0 となる前に弦より離れる。なお、撥弦体と弦が接触しつつ移動する間に生じる摩擦力については無視している。

この撥弦モデルは次式の様に表される。

$$M_H \frac{d^2 \eta}{dt^2} = -F_H(t),$$

$$F_H(t) = \begin{cases} K|\eta(t) - y(x_0, t)|^p, & \eta(t) \geq y(x_0, t), t < t_f \\ 0, & \text{others} \end{cases}$$

ここで、 η : ハンマーの変位、 $F_H(t)$: 弦に加わる力、 M_H : ハンマーの質量、 t_f : 撥弦体と弦との接触時間 (力の継続長) である。 K, p は実験的に決められる定数である。また、 $f(x, x_0, t)$ と $F_H(t)$ の関係は次式で与えられる。

$$f(x, x_0, t) = \frac{F_H(t)g(x, x_0)}{\mu \int_{x_0-\delta x}^{x_0+\delta x} g(x, x_0) dx}$$

$2\delta x$: ハンマーの幅、 $g(x, x_0)$: 力の分布を与える関数である。

境界条件は両端支持とし、初期条件として $g(x_0, x)$, x_0, V_{H0} を与える。

有限差分法を用いてこれらの方程式を解く。そして、合成パラメータとして物理定数を指定し、変位等を再帰的に計算する。

謝辞

本研究を進める機会を与えて下さった当所情報科学研究部 石井健一郎部長、有益な議論を頂いた聴覚研究グループ植松尚研究員、および日頃討論頂く音表現研究グループの皆様に感謝します。

参考文献

- [1] 小坂 直敏, “Sinusoidal model による音色の補間”, 電子情報通信学会技術報告 SP95-130, pp9-16(1996).
- [2] 小坂 直敏, “非定常信号を含む楽音の音色補間”, 日本音響学会平成9年度春季研究発表会講演論文集, 2-9-11, pp657-658(1997).
- [3] 引地 孝文, 小坂 直敏, “打弦、撥弦、および弾性体振動のモルフィングの検討”, 電子情報通信学会技術報告 SP96-111, pp.23-28(1997).
- [4] 難波 精一郎, “音色の定義を巡って”, 日本音響学会誌, 49 巻 11 号, pp.823-831(1993).

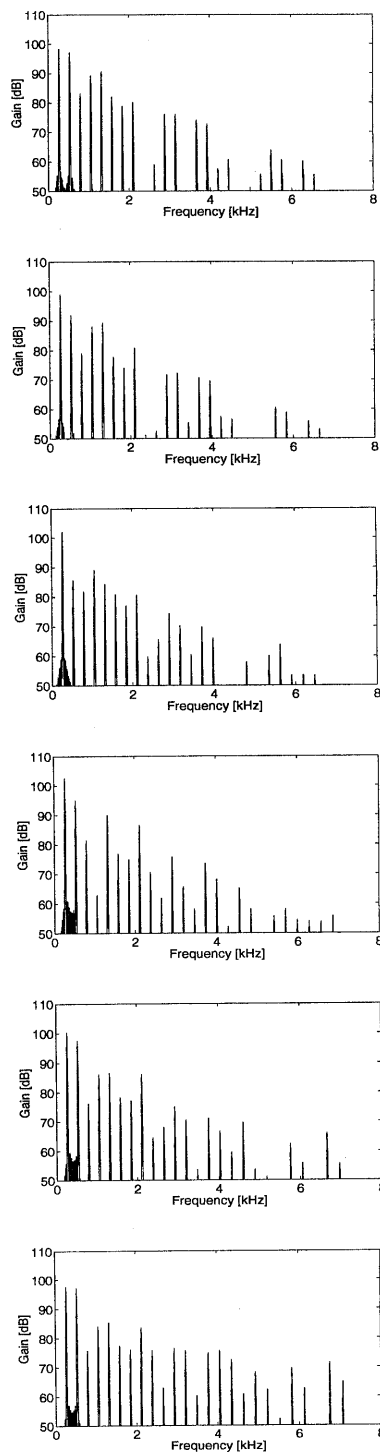


図 14: 評価試験に用いた合成音の FFT スペクトルの一例 ($\alpha = 0.0(\text{NS}), 0.2, 0.4, 0.6, 0.8, 1.0(\text{PI})$)