

梵鐘音の特徴に関する基礎的考察 II

大 島 静 夫・伊 藤 禎 紀^{*1}

Basic Consideration on Characteristics of Sound of Japanese Temple Bell II

Shizuo OHSHIMA and Yoshiki ITOH

(1999年11月25日受理)

Temple bells have already been the subject of many studies. Our first report analyzed the distribution in the time-frequency domain of the bell's sound using the STFT (short-time FFT) method. The results showed that the partial tones that make up the sound of the temple bell could be approximated as the product of an exponential attenuation term and a sine wave term.

In our second report, we examined methods of reproducing ten temple bell sounds made up of parts up to the third partial tone. By estimating the exponential attenuation term and beat frequency of each part by STFT and by estimating the central frequency of each partial tone and its initial phase by entire-interval FFT, we found it was possible to digitally reproduce the sound of a temple bell as the sum of their products. The energy error between the original tone and the composite tone was 5% or less in the STFT region.

1. はじめに

梵鐘に関しては、既に多くの研究がなされている。その中で我々は、梵鐘音をできるだけ少ないパラメータで、合成音として再現し、各パラメータのもつ意味を把握し、等価回路的な表現ができないかを検討している。第一報として、STFT (Short time FFT) の手法を使用し、梵鐘音が時間一周波数領域でどのように分布するかを示し、その結果、梵鐘音を構成する主要な部分音は最大でも7次程度までであり、各部分音の包絡線は指数減衰項と正弦波項の積で近似できることを示した。

第二報では、CDに録音されている梵鐘音のうち、ほぼ第三部分音までで構成される10梵鐘音に的を絞り、その音を再現する手法について検討を行った。その結果、STFTで各部分の指数減衰項と正弦波項を推定し、全区間FFTで各部分音の中心周波数とその初期の位相を推定し、その積和として梵鐘音をほぼ合成できることがわかった。以下に、本手法による合成音の再現法およびその評価を示す。

2. 解析方法

2. 1 データの収録

梵鐘の解析に関する文献によると、鳴り始めから約1秒間の梵鐘音は、打音部といわれ、多数の部分音と雑音成分から成る複雑な音であり、鳴り始めから0.8秒以降の梵鐘音を解析対象としていることから、ここでも同様の条件でデータを解析することにした。

CDからのデータ取り出しには、アプリケーションソフト (Wave Studio) を用い、視察により梵鐘の立ち上がり部分に近いゼロクロス点を仮のスタート点とし、それから0.8秒間のデータを削除し、更にその点から65536ポイント (約6秒間) のデータを切り出し、解析データとした。サンプリング量子化量は16[bit]、サンプリング周波数は、主に低周波数に解析が限定されることから11025[Hz]とした。以降、梵鐘の番号はCDの録音順にBON01~BON84とする。

^{*1} 秋田高専専攻科学生

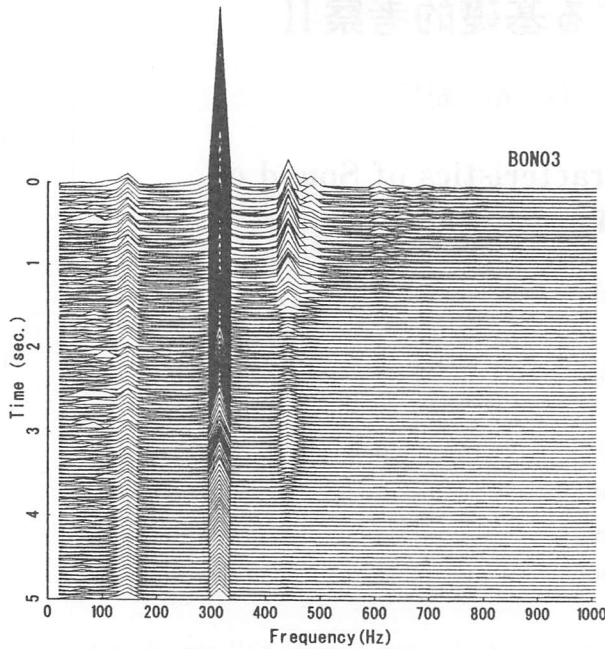


図1 梵鐘のスペクトル分布例

2.2 短区間 FFT 分析

まず予備的な解析として、84個の梵鐘音の STFT 分析を行ない、そのスペクトル分布を調べた。サンプリングポイント数は512点としたので、STFT 解析の基本周波数は、 $f_0 = 11025/512 = 21.53$ [Hz]、約6秒間(65536ポイント)の解析フレーム数は128となる。またデータの切り出しには、フレーム内でのデータの変動幅が小さいため矩形窓を用いた。

一般にスペクトル図において振幅は、その詳細を見定めるためにパワースペクトルを取るが、ここでは概略の振幅分布を見定めることが目的なので、線形軸のほうが、その特徴が明確になると考え、線形な軸を採用した。図1に STFT の解析例として BONG03のスペクトル分布図を示す。

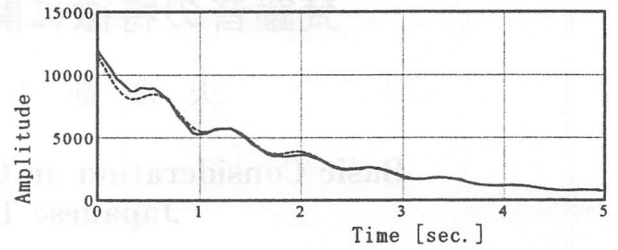
図1に示す、打音部以降のスペクトルは、比較的単純な構造であり、また基音から比較的低次の部分音の和で再現できることがわかる。そこでこの梵鐘音を(1)式で近似表現する。

$$b(t) = \sum_{n=1}^3 b_n(t) \sin(2\pi f_{n0}t + \theta_{n0}) + err \quad (1)$$

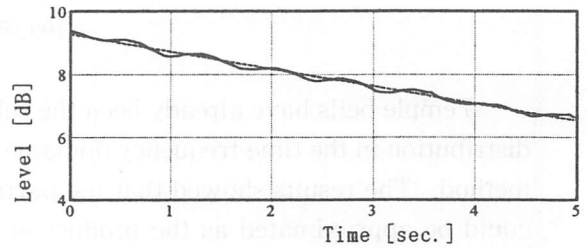
ここで $b(t)$ は実際の梵鐘音の振幅、 $b_n(t)$ は各部分音の振幅、 f_{n0} は部分音の中心周波数、 θ_{n0} は部分音の初期位相、 err は全体の誤差である。そしてここでは基音も含め、第三部分音までの和で表現することより $n = 3$ とした。

2.3 部分音の解析

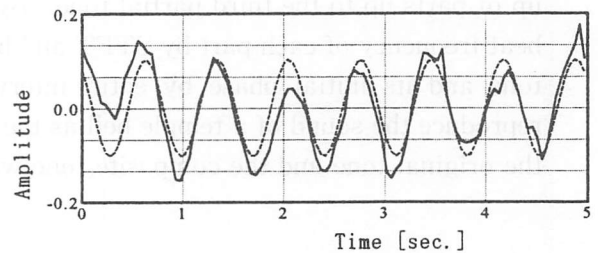
梵鐘の STFT 解析で得られたデータから各部分



(a) BONG03 部分音と近似波形



(b) (a)の自然対数波形と近似波形



(c) (b)の残差と残差の近似波形

図2 部分音のパラメータの抽出例

音の中心周波数の時間波形の解析を試みた。ここでは、BONG03の図1における300[Hz]付近の部分音を解析例として示す。

図2(a)から、波形は指数関数的な減衰項および正弦波的なうねりから構成されていることが推定できる。そこで部分音 $b_n(t)$ を(2)式で近似できるものとする。

$$b_n(t) = a_n \exp(-b_n t) \{ a_{bn} \sin(2\pi f_{bn} t + \theta_{bn}) + d_n \} + err \quad (2)$$

ここで a_n は第 n 部分音の初期振幅、 b_n は減衰係数、 a_{bn} は n 周波数の振幅、 f_{bn} は n 周波数、 θ_{bn} は n 周波数の位相角、 d_n は直流分とする。

まず(2)式の指数減衰項を求めるために、(2)式を自然対数をとると(2)式は

$$\log_e b_n(t) = \log_e a_n - b_n t + \log_e(err') \quad (3)$$

となる。ただし err' には、正弦波項の変動も近似における誤差として含めている。対数化した原波形を最小二乗近似法を用い、 $y = m_a x + n_a$ と直線近似すると、傾き m_a および切片 n_a より

梵鐘音の特徴に関する基礎的考察 II

$$b_n = -m_a, \quad a_n = \exp(n_a) \quad (4)$$

と定まる。図 2 (b) に対数化した部分音の原波形と (4) の係数による直線近似式を示す。

(4) で求めた $a_n \exp(-b_n t)$ で、(2) 式を割ると

$$\frac{b_n(t)}{a_n \exp(-b_n t)} = a_{bn} \sin(2\pi f_{bn} t + \theta_{bn}) + d_n + err'' \quad (5)$$

となる。この (5) 式の平均値を求めると、 t が十分長いとき

$$\sum_t = a_{bn} \sin(2\pi f_{bn} t + \theta_{bn}) \doteq 0, \quad \sum_t err'' \doteq 0$$

と考えられるので、(5) 式の平均値より d_n が定まる。今フレーム数を m とすると

$$\frac{1}{m} \sum_{t=1}^m \left(\frac{b_n(t)}{a_n \exp(-b_n t)} \right) \doteq d_n \quad (6)$$

となる。この直流分も除くと (5) 式は

$$\frac{b_n(t)}{a_n \exp(-b_n t)} - d_n = a_{bn} \sin(2\pi f_{bn} t + \theta_{bn}) + err'' \quad (7)$$

となる。(7) 式までの処理を施した部分音の原波形を図 2 (c) に示す。

この波形を $a_{bn} \sin(2\pi f_{bn} t + \theta_{bn})$ に近似すると、図 2 (c) の点線の波形となる。第一報では、更にこの残差も正弦波近似していたが、より簡単にするために本報ではここで近似を打ち切ることとした。

次に、ここまでのパラメータを使用した部分音の合成波形を $s_n(t)$ とすると

$$s_n(t) = a_n \exp(-b_n t) \{ a_{bn} \sin(2\pi f_{bn} t + \theta_{bn}) + d_n \} \quad (8)$$

となる。ここで部分音の原波形と合成波形のエネルギー誤差を (9) 式と定めると、図 3 (a) の場合では 0.43 [%] となった。

$$err_n = \frac{\sum (b_n(t) - s_n(t))^2}{\sum b_n^2(t)} \times 100 [\%] \quad (9)$$

また 10 梵鐘の最大振幅をもつ部分音の (9) 式による誤差は、0.008 [%] ~ 1.361 [%] であり、その平均値は 0.355 [%] であった。

2. 4 全区間 FFT 解析

STFT の解析結果を用いても、梵鐘音の合成は可能であるが、STFT の結果を利用した場合、周波数分解能が 21.53 [Hz] と粗く、正確な中心周波数が定まらない。そこで各部分音の中心周波数 f_{no} をより正確に求めるために 65536 ポイントの全区間でハミング窓を掛けた上で FFT 分析を行った。この場合その周波数分解能は $11025/65536 = 0.1682$ [Hz] と

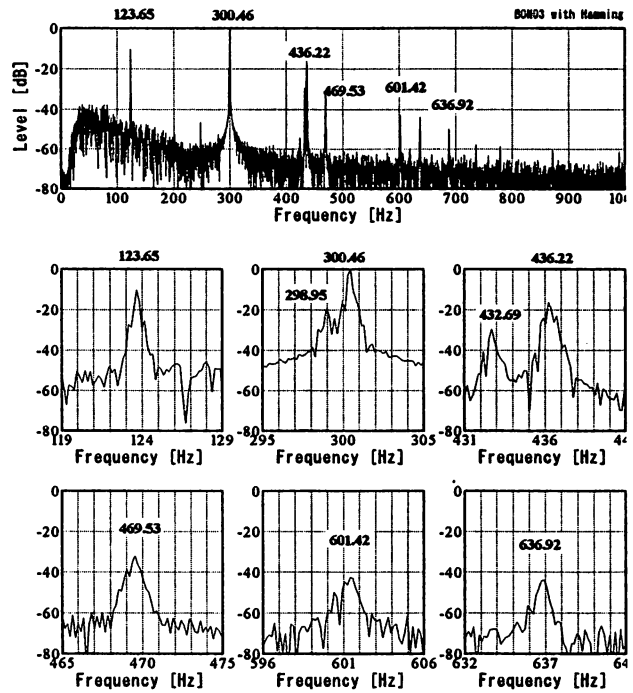


図 3 全区間 FFT の解析例

る。BON03の全区間 FFT の結果および部分音中心周波数付近のスペクトルの拡大図を図 3 に示す。

図 3 より、 $f_{10} = 123.65$, $f_{20} = 300.46$, $f_{30} = 436.22$ [Hz] と定めた。また各部分音の中心周波数の初期位相 θ_{no} は、FFT 解析の実部 X_{no} 、虚部 Y_{no} の逆正接として (10) 式で定めた。

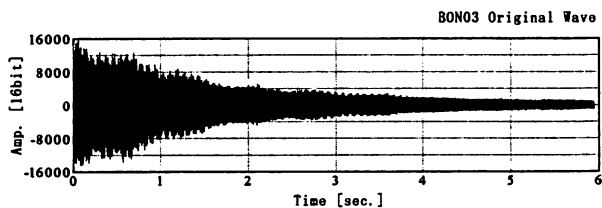
$$\theta_{no} = \tan^{-1} \frac{Y_{no}}{X_{no}} \quad (10)$$

ここでの問題としては、部分音が近接した 2 つの周波数で構成される場合、どのように取り扱うかが上げられる。すなわち、振幅の大きい周波数を採用するか、2 つの周波数の平均値を使うか、2 つの周波数を合成波形に組み入れるかである。本報告ではより合成波形に影響を与える振幅の大きい周波数のみを選択したが、この部分については更に検討を加える必要があると思われる。

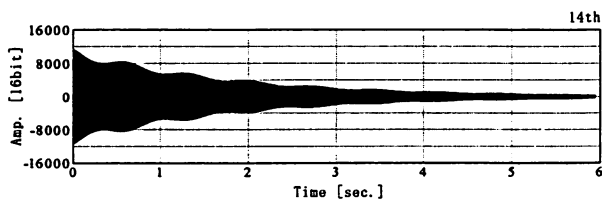
2. 5 合成音の作成とその評価

ここまでで (1) 式の各部分音を構成するパラメータが全て定まったことになる。そこでエネルギーの大きい順に各部分音を順次積算した時間波形を図 4 に示す。図 4 において (a) は原波形、(b)(c) は各部分音の振幅順の順次累積合成波形、(d) は第三部分音までの合成音波形である。原波形と比較しても、微細な部分も含め、近似できていることが分かる。

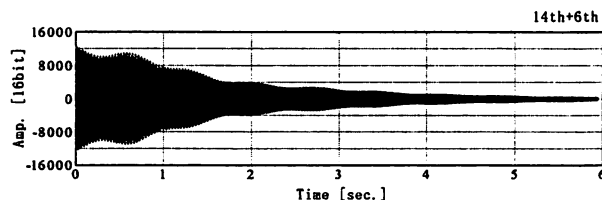
同様に他の梵鐘の原波形と合成波形の比較図を図



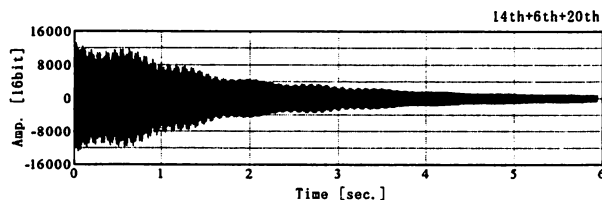
(a) 原波形



(b) 第二部分音



(c) 第一と第二部分音の累積波形



(d) 第三部分音までの累積波形

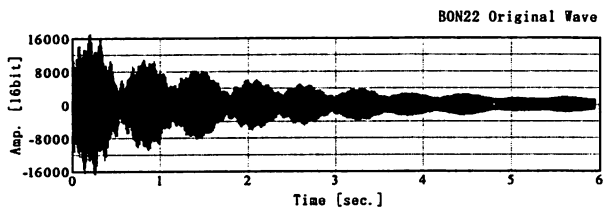
図 4 原波形と合成波形

5に示す。

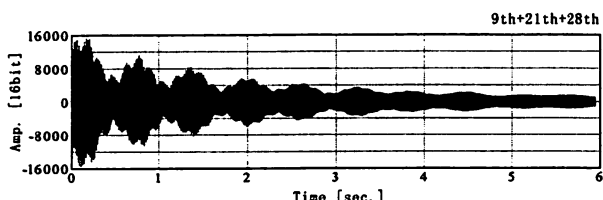
まだ合成音の聴感上の評価は行っていないが、高次の部分音を含む立ち上がり部分を除き、ほぼ同一音として捉えることができ、また梵鐘のうなり音も差異無く聞き取ることができる。

ただし、時間領域で原波形と合成波形の差分を誤差として評価した場合、原音の部分音の中心周波数が時間的に変動するため、周期的な逆相状態を作り出し、時間領域での誤差が大きくなる。そこで合成音をさらにSTFTし、原波形との時間-周波数領域での違いを誤差と定めた。

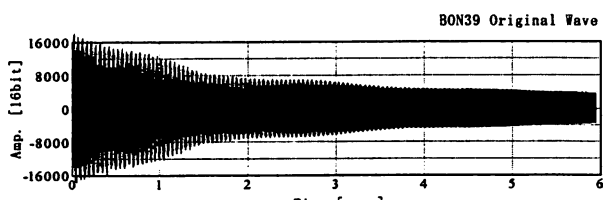
原波形および合成波形をSTFT処理すると、時間tに関し128フレーム、各フレームの周波数項は256次までのデータで表現できる。そこで、両波形を(11)



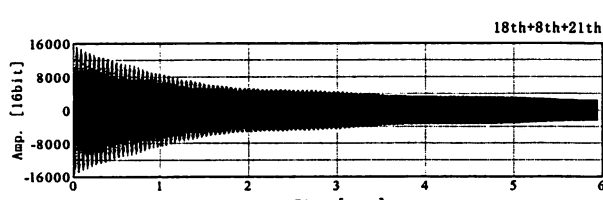
(a) BON22 原波形



(b) BON22 合成波形



(c) BON39 原波形



(d) BON39 合成波形

図 5 原波形と合成波形の比較

のように離散データとして表現する。

$$STFT(original) = b[t, f]$$

$$STFT(composit) = s[t, f] \tag{11}$$

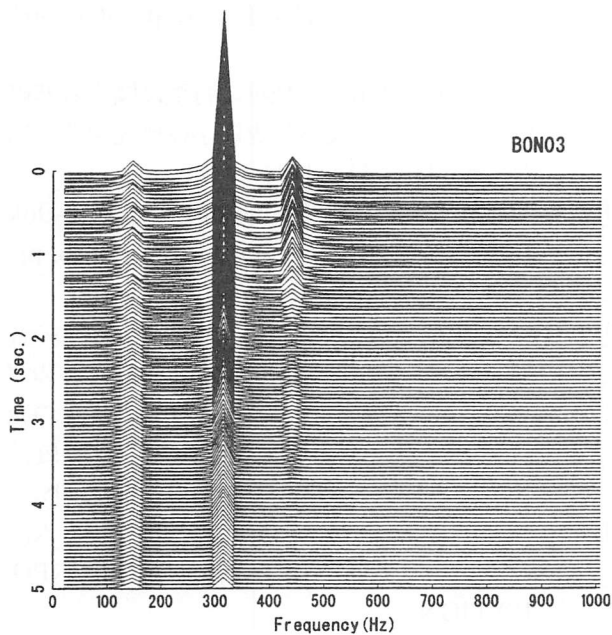
ただし、 $t=1, 2, \dots, 128$, $f=1, 2, \dots, 256$

これより、両式の差分のエネルギー誤差を(12)式のように定めた。

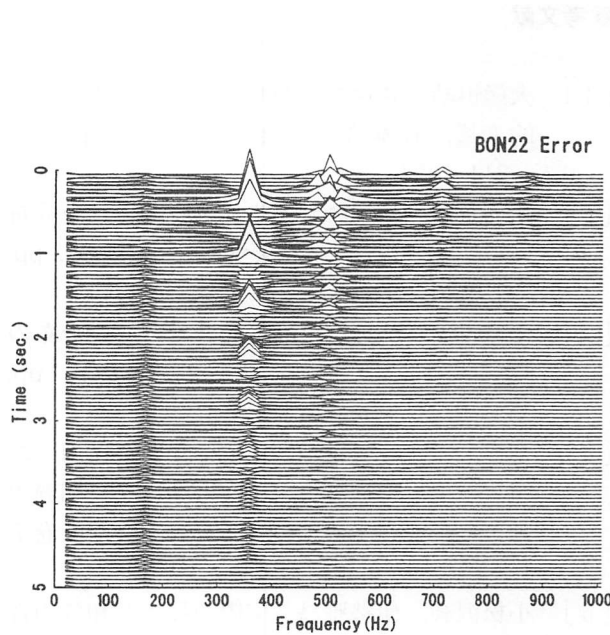
$$err = \frac{\sum \sum (b[t, f] - s[t, f])^2}{\sum \sum b^2[t, f]} \times 100[\%] \tag{12}$$

図 6(a)に合成波形のSTFTのスペクトル、図 6(b)に原波形と合成波形との残差のスペクトル図を示す。

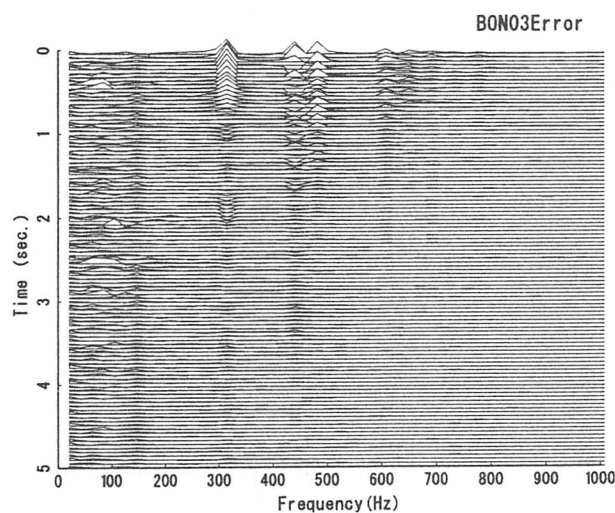
図から各部分音の主要なピークが取り除かれていることがわかる。BON03の場合(12)式による誤差は



(a) 合成音のスペクトル

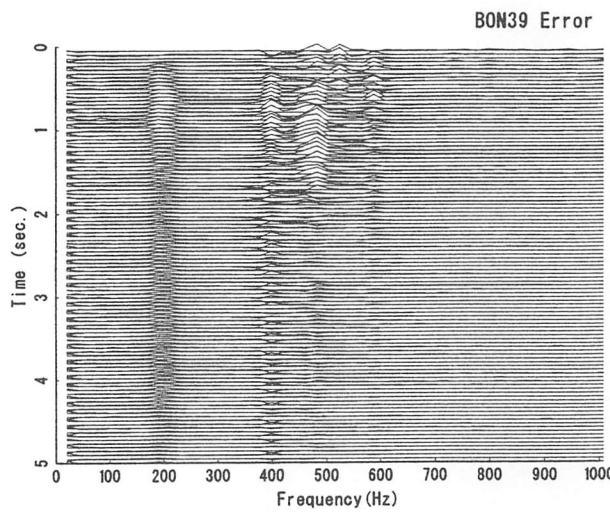


(a) BON22 の原音と合成音の誤差



(b) 原音と合成音の誤差

図 6 合成音と原音との誤差



(b) BON39 の原音と合成音の誤差

図 7 他の梵鐘の合成音と原音との誤差

1.80[%]であった。

図 7 に同様に、BON22、と BON39 の原波形と合成波形の時間一周波数領域における誤差を示す。

3. 結 論

パラメータを指定するだけで、梵鐘の音色を再現することを目的に、第二段階として、第三部分音までで構成される梵鐘音の合成音の作成とその評価を試みた。

その結果、STFT で各部分音の指数減衰項とうなり周波数を推定し、全区間 FFT で各部分音の中心

周波数とその初期位相を推定し、その積和として梵鐘音を合成した場合、STFT 領域でのエネルギー誤差は BONO3 の場合で約 1.80[%] となった。また、ここで取り扱った 10 梵鐘では 0.51[%] ~ 4.73[%] であり、全体の平均値は 2.55[%] となった。

このことから、第三部分音までの構成で近似できる梵鐘の場合、各部分音を 8 パラメータで合成できるので、全体として 24 パラメータで合成できることになり、またエネルギー誤差が 5 [%] 以下であることより、梵鐘音の 95 [%] 以上を再現できたことになる。

今後はこのパラメータの持つ意味を把握し、等価回路的な表現法についても検討したい。

参考文献

- [1] 大熊恒靖：梵鐘の音の減衰時間に関する時代的変遷，日本音響学会誌53巻3号，pp. 208～214, 1997
- [2] 栗原正次：梵鐘における部分音の振動数分布の時代的変遷，日本音響学会誌37巻12号，pp. 599～605, 1998
- [3] 栗原正次：梵鐘における部分音の強度分布の時代的変遷，日本音響学会誌37巻12号，pp. 606～612, 1981
- [4] 小松沢昶，伊勢雅昌：撞木の代わりに用いた材質の塊が和鐘の音色に及ぼす影響（和鐘の音色の研究 その1），日本音響学会誌32巻7号，pp. 413～419, 1976
- [5] 小松沢昶，伊勢雅昌：実際の撞木が和鐘の音色に及ぼす影響（和鐘の音色の研究 その2），日本音響学会誌32巻8号，pp. 465～469, 1976
- [6] 大島静夫：秋田県の梵鐘の特徴に関する基礎的考察，秋田工業高等専門学校研究紀要第33号，pp. 110～115, 1998
- [7] Yang-Hann Kim : Sound of King Song-Dok Bell, J. Acoust. Soc. Jpn. (E) 19, Vol. 5, pp. 305～313, 1998
- [8] 降旗建治，柳沢武三郎，池田雅昭：木魚の音響特性，信学技報 EA97-46, pp. 15～22, 1997
- [9] 春原政浩，柳沢武三郎，降旗建治：木魚の発音メカニズムおよび合成音を用いた音質評価，信学技報 EA98-104, pp. 21～26, 1998
- [10] 熊谷恭孝：秋田の梵鐘～梵鐘レクイエム，Compact Disc (YKCD-1001), OZ CORPORATION