MATHEMATICAL MODELING OF HARMONIC-TIMBRE STRUCTURE WITH MULTI-BETA-DISTRIBUTION

Toru Nakashika

Kobe University
Graduate School of Engineering, Japan
nakashika@me.cs.scitec.kobe-u.ac.jp

Tetsuya Takiguchi, Yasuo Ariki

Kobe University
Graduate School of Engineering, Japan
takigu@kobe-u.ac.jp ariki@kobe-u.ac.jp

ABSTRACT

Recently, a large amount of signal processing technology research on applications associated with music is being carried out. Sound synthesis, in particular, is one of the most interesting research themes. In this paper we propose a new approach to mathematically modeling harmonic-timbre structure with multi-beta-distribution (MBD). This probabilistic distribution has the advantage of enabling one to easily express varied timbre-structure with only a few parameters. We will define MBD itself, and present a method of estimating MBD parameters. Some experimental results are presented to discuss the performance of this method.

Index Terms— music, synthesis, modeling, timbre, multi-beta-distribution, MBD

1. INTRODUCTION

As computers are able to synthesize timbre sounds, parametrically modeling timbre structure has become common. Related research on mathematically expressing timbre structure are discussed in [1] using a first sinusoidal representation and in [2] using a Partial Damped & Delayed Sinusoidal model. In another study, S. L .Groux proposed a system that allows the control of an additive sound synthesis model intuitively from sonic features [3].

This paper proposes a new approach to extract timbre features with multi-beta-distribution (MBD). MBD is based on beta function, which can express the varied shapes of distribution by changing the value of two parameters α and β. We, therefore, can express diverse, 2-dimensional structures, such as timbre shape, using a MBD model. Once we model timbre structures, we can store them in the database of MBD parameters, and we can extract timbre sound whenever we like. Moreover, it is supposed that it be efficient for automatic music transcription or source separation [4][5][6], since MBD-modeled timbre structure is represented as a mathematical formula.
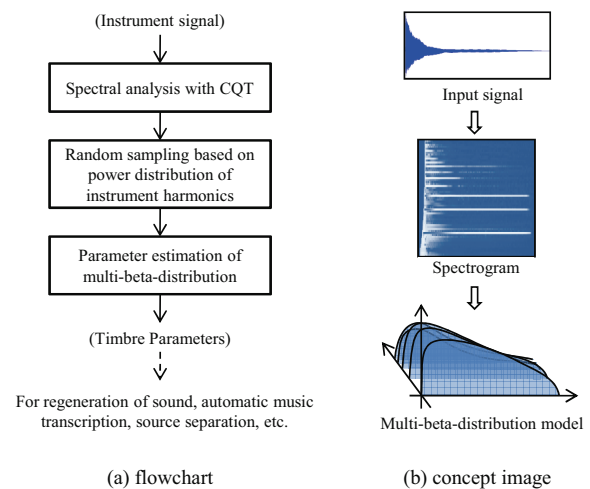

Fig. 1. Overview of modeling timbre structure with multi-beta-distribution

2. OVERVIEW OF THE PROPOSED METHOD

In this section, we give an overview of the proposed method with which we can express timbre shapes mathematically. The outline of the timbre modeling technique we propose is shown in Figure 1, where (a) shows a flowchart of the proposed method and (b) a concept image of processing.

The pitch and duration of the input sound data used for experimental estimation of MBD parameters are normalized beforehand. In the first stage, the source is analyzed with Const-Q-Transform (CQT) [7]. As a result of this analysis, we can read out each intensity distribution corresponding to fundamental pitch and to the other harmonics. In the proposed method, since this power distribution is taken out discretely in each overtone, the power distribution can be formulated as a set of beta distributions, which we call multi-beta-distribution (MBD), as shown in Figure 1 (b) below. If we are able to acquire the MBD parameters, we can easily use it for generating timbre sound, for automatic music transcription or for sound separation.
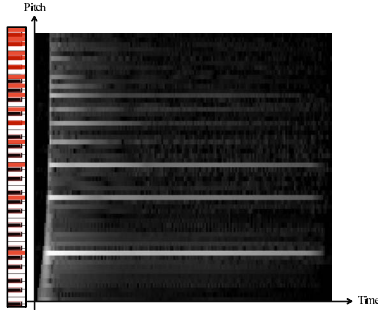
Fig. 2. Spectral analysis of piano tone with CQT

3. SPECTRAL ANALYSIS USING CQT

3.1. Analysis of Musical Sound with CQT

The Constant Q transform (CQT) was derived by Brown as a means of creating a log-frequency resolution spectrogram [7]. This has considerable advantages for the analysis of musical signals because the frequency resolution can be set to match that of the equal tempered scale used in western music, where the frequencies are geometrically spaced, as opposed to the linear spacing that occurs in the discrete Fourier transform. Therefore, we used CQT for analyzing input signals in this paper.

CQT is defined as:

$$X[k] = \frac{1}{N_k} \sum_{n=0}^{N_k-1} W_{N_k}[n]\, x[n] e^{\frac{-j2\pi Qn}{N_k}} \qquad (1)$$

where

$$N_k = Q\frac{f_s}{f_k} \qquad (2)$$

$$Q = \left(2^{\frac{1}{b}} - 1\right)^{-1} \qquad (3)$$

$$f_k = f_0 2^{\frac{k}{b}} \quad (k = 0, 1, ...) \qquad (4)$$

denote a size of the k-th window, a constant, and the frequency we want to analyze, respectively. We let b be a number of bin in an octave, f_0 be the initial frequency and f_s be the sampling frequency.

$x[n]$ is an input signal on the time-domain, and W_{N_k} is a window function which has a size of N_k.

3.2. Results of Spectral Analysis

Figure 2 shows the result of analyzing piano sound with CQT.

This figure shows the analysis of a single tone A4 on the piano, and the red parts on the left indicate overtones of A4. Looking at this figure, only the harmonic part is analyzed with CQT; the intensity of the other portions is hardly shown. This implies that we can extract intensity distributions corresponding to each harmonic.

4. FORMULATION OF MULTI-BETA-DISTRIBUTION

In this section, we discuss the multi-beta-distribution (MBD) proposed in this paper.

4.1. Overview and Definition of Multi-beta-distribution

4.1.1. Overview

We consider MBD to be a distribution in which two or more beta distributions overlap another dimension, as shown in Figure 3, and when we model a timbre structure using MBD, we call it a Multi-beta-model (MBM).

Beta distribution, which is represented by the following equation (5), is a probability density function that can express very diverse distribution shapes by changing the value of two parameters α and β.

$$p(t; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} t^{\alpha-1}(1-t)^{\beta-1} \qquad (5)$$

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt \qquad (6)$$

$$= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \qquad (7)$$

where $\Gamma(x)$ denotes gamma function. We propose that the distribution that arranges the beta distributions to another dimension axis is unquestionably multi-beta-distribution. By using this distribution MBD, we can express 2-dimensional shapes easily.

4.1.2. Definition

Now, we will define multi-beta-distribution. MBD is the distribution that has a structure in which each beta distribution exists discretely in the axis. Each distribution can be expressed, with the i-th beta distribution being expressed as $p(t; \alpha_i, \beta_i)$. In addition, all distributions are weighted by the energy ratio c_i and do not exist mutually independent.

The multi-beta-distribution is defined as follows (8).

$$q(x, t; \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{c}) \triangleq \sum_n \delta_{x,(\log n)-1} c_n p(t; \alpha_n, \beta_n) \qquad (8)$$

where

$$\sum_n c_n = 1, \quad \forall n,\, \alpha_n, \beta_n > 0, \quad \delta_{i,j} = \begin{cases} 1 & (i = j) \\ 0 & otherwise \end{cases} \qquad (9)$$

are satisfied, and x, t and n indicate the variables for log-frequency, time, and the index of harmonics, respectively.

In this paper, we use MBD to model each timbre structure. We suppose that the feature of each instrument can be distinguished from others, so the timbre structure modeled with MBD is invariable for absolute pitch. In other words, each
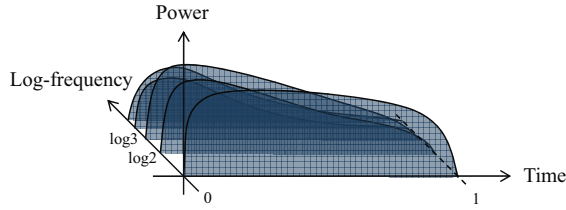
Fig. 3. Entire picture of multi-beta-distribution

instrument structure requires its own MBD, and with respect to a certain timbre, all the harmonics have the same shape as that of fundamental tone.

4.2. Parameter Estimation

4.2.1. Estimation of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$

Parameters of beta distribution α and β can be estimated by using the method of maximum likelihood [8]. Assuming we get K samples $X_1 \leq \cdots \leq X_K$ from a certain beta distribution $p(t; \alpha, \beta)$, the log likelihood function is shown by expression (10).

$$
\begin{aligned}
l(\alpha, \beta) &= \log \prod_{i=1}^{k} p(X_i) \\
&= -k \log [B(\alpha, \beta)] + (\alpha - 1)\left[\sum_{i=1}^{k} \log X_i\right] \\
&\quad + (\beta - 1)\left[\sum_{i=1}^{k} \log(1 - X_i)\right]
\end{aligned}
\tag{10}
$$

Solving $\frac{\partial l}{\partial \alpha} = 0, \frac{\partial l}{\partial \beta} = 0$, we can obtain

$$
\left\{
\begin{aligned}
\Psi(\alpha) - \Psi(\alpha + \beta) &= \frac{1}{k}\sum_{i=1}^{k} \log X_i \tag{11} \\
\Psi(\beta) - \Psi(\alpha + \beta) &= \frac{1}{k}\sum_{i=1}^{k} \log(1 - X_i) \tag{12}
\end{aligned}
\right.
$$

where, $\Psi(x) = \frac{d}{dx}\log\Gamma(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ indicates the digamma function. Since the right side of the expression (11) is already-known, α and β that satisfy both expressions simultaneously can be approximately solved.

The above discussion is the same for each parameter of MBD $\boldsymbol{\alpha} \in \boldsymbol{R}^n, \boldsymbol{\beta} \in \boldsymbol{R}^n$.

4.2.2. Estimation of $\boldsymbol{c}$

The weight parameters vector between two or more beta distributions is represented by $\boldsymbol{c}$. It is natural to think that $\boldsymbol{c}$ can be expressed as the ratio of the fundamental tone to the average intensity of each of the harmonics in a timbre structure.

From the sound signal for training, the parameter $\boldsymbol{c} \ni c_n$ can thus be obtained as given by (15).

$$
m_n = \frac{1}{N}\sum_{i=0}^{N-1} g_n[i] \tag{13}
$$

$$
K_n = \frac{m_n}{m_1} = \frac{\sum_{i=0}^{N-1} g_n[i]}{\sum_{i=0}^{N-1} g_1[i]} \tag{14}
$$

$$
c_n = \frac{K_n}{\sum_n K_n} \tag{15}
$$

$g_n[i]$ means the intensity distribution of the n-th harmonic. m_n indicates the average of $g_n[i]$.

5. EXPERIMENTAL RESULTS AND DISCUSSION

5.1. Experimental Setup

We tested our approach experimentally to evaluate it, modeling timbre structure with MBD, from an audio signal. The input signal of piano sound performed in about 3.9 ms was recorded from MIDI sound source by the sampling rate 22.05kHz. Letting the number of bin be 36, the number of harmonics be 8, the minimum frequency be 220Hz (A3), and the maximum frequency be 7040Hz (A8), we analyzed the input signal through CQT, estimated the MBD-parameters, and then verified the performance of the proposed method experimentally.

5.2. Experimental Results

The experimental results of modeling the timbre structure of the piano using the proposed method is shown in Figure 4. The left portion of Figure 4 represents the original timbre structure of a test data, and the right portion a structure modeled by MBD. From left to right in both figures, the distributions of fundamental and second tone, etc. are mutually displayed.

In addition, to increase visibility, 2-dimensional plots carried out in the same way are shown in Figure 5, where (a), (b), (c) and (d) represent the results for the fundamental, 8th, 5th and 3rd harmonic, respectively. The vertical axis of the graph indicates time, and the horizontal axis intensity. In Table 1, we list the results of the estimated parameters of MBD. 'h1', 'h2', etc. refer to the result for fundamental, second harmonic, etc.

5.3. Discussion

According to the results shown in Figure 4, we can deduce that the MBD approach proposed in this paper has the ability to express features of a timbre structure, which contain the intensity-ratio between fundamental and each of the harmonics and the decay of the whole intensity as time passes, etc. The features of piano timbre structure are as follows: the average intensity of the fundamental tends to be higher than that
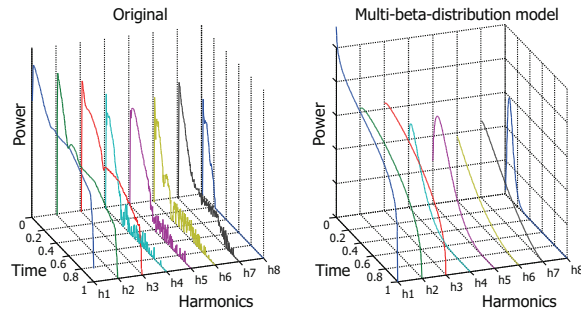
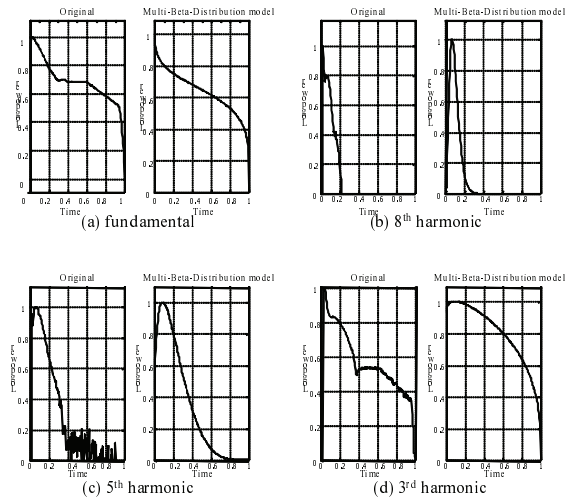**Fig. 4**. Result of Modeling Timbre Structure with MBD

Fig. 5. Modeling Results (2D View)

Table 1. Estimated parameters of multi-beta-distribution

	• •	• •	c
h1	0.9340	1.1868	0.301137
h2	1.0227	1.3798	0.204969
h3	1.0172	1.3889	0.203554
h4	1.2040	5.7828	0.053784
h5	1.3916	5.1418	0.075078
h6	1.0109	3.2020	0.051976
h7	1.0282	2.5836	0.080684
h8	2.9105	28.0358	0.028818

6. CONCLUSION

In this paper, we proposed a new approach for expression of timbre structure, which is based on multi-beta-distribution (MBD). We invented a formulation and a parameter estimation of MBD, and also generated a timbre sound from a trained MBD model. We verified experimentally that the MBM model is sufficiently suited to expressing timbre structure.

In the future, we would like to expand our research on automatic music transcription, timbre separation, and other related themes.

7. REFERENCES

[1] R. J. McAulay and T. F. Quatieri, "Speech Analysis&Synthesis Based on a Sinusoidal Representation," *IEEE Trans. on Acoust. Speech and Signal Proc.*, vol. 34, pp. 744–754, 1986.

[2] Remy Boyer and Karim Abed-Meraim, "Efficient Parametric Modeling for Audio Transients," *Proceedings of the 5 th International Conference on Digital Audio Effects*, pp. 26–28, September 2002.

[3] Sylvain Le Groux *et al.*, "Perceptsynth: Mapping Perceptual Musical Features to Sound Synthesis Parameters," *IEEE ICASSP 2008*, pp.125–128, April 2008.

[4] H. Kameoka *et al.*, "A Multipitch-Analyzer Based on Harmonic Temporal Structured Clustering," *IEEE Trans. on Audio, Speech and Language Processing*, vol. 15, pp. 982–994, March 2007.

[5] K. Miyamoto *et al.*, "Harmonic-Temporal-Timbral Clustering (HTTC) for the Analysis of Multi-instrument Polyphonic Music Signals," *IEEE ICASSP 2008*, pp. 113–116, April 2008.

[6] K. Miyamoto *et al.*, "Probabilistic Approach to Automatic Music Transcription from Audio Signals," *IEEE ICASSP 2007*, vol. 2, pp. 697–700, April 2007.

[7] Brown, J.C., "Calculation of a Constant Q Spectral Transform," *J. Acoust. Soc. Am. 89*, pp. 425–434, 1991.

[8] R. Gnanadesikan *et al.*, "Maximum Likelihood Estimation of the Parameters of the Beta Distribution from Smallest Order Statistics," *Technometrics 9*, pp. 607–620, 1967.

of the other harmonics and the higher the index of harmonics, the more the average intensity of the harmonics declines. In addition, the intensity of each harmonic tends to decline once you go beyond the attack time.

As shown in Figure 5 (a) and 5 (b), each beta distribution of MBD can be approximately expressed with respect to intensity-peak time and intensity-damped curve, etc.

Looking at Figure 5 (c), you can see that some noise exists in the lower intensity region on the original distribution. However, it does not occur on the modeled distribution. This means that MBD has robustness against noise.

As shown in Figure 5 (d), there was an example in which MBD could not accurately estimate the original distribution. This is because beta distribution cannot express a structure that has two peaks, as shown above.

Table 1 shows the parameters of MBD. *c* means a mixing coefficient (the ratio of intensity distribution between harmonic overtones), and we can numerically see that MBD has the ability to express the features of a piano, where the intensity is decreasing as the index of harmonics progresses except h4 and h7.