Evaluation System for Groove Feelings Evoked by Drum Sounds

Hiroto Kawata¹, Takeshi Hori¹, Kazuyuki Nakamura^{1,2}

¹Graduate School of Advanced Mathematical Sciences, Meiji University
4-21-1 Nakano, Nakano-ku, Tokyo 164-8525 Japan

²JST, PRESTO

4-1-8 Honcho, Kawaguchi, Saitama, 332-0012, Japan

E-mail: {cs181004, hori, knaka}@meiji.ac.jp

Abstract

In this study, we define stochastic models for tight and loose groove feelings and propose an evaluation system for the groove feelings evoked by drum sounds. There are two main types of groove feelings, tight and loose, and they are known among professional musicians and music enthusiasts. These groove models are defined as a bivariate normal distribution with deviations for snare and bass drums as random variables. As an application of these models, we propose our evaluation system, in which a monaural, polyphonic drum set recordings is input and evaluations regarding tight and loose are output. We conducted an evaluation experiment and showed that this system correctly identifies tight and loose groove feelings.

1 Introduction

Groove feelings are considered to be an expression of a musical performance arising from rhythm[1]. Kawase and Eguchi[2] conducted a subjective survey and showed that the groove feeling is close to the Japanese word “*nori*”. In particular, in a drum performance, there are two main groove feelings, tight and loose, and they are known among professional drum players and music enthusiasts. These groove feelings express the auditory impression of a drum performance.

Groove feelings are abstract concepts. Although it is known that to play drums with groove feelings hit-points of snare and bass drums must be placed earlier/later than score-based hit-points, this groove description is subjective because it is given by an oral or written instruction and they are not defined quantitatively. For this reason, the definition of groove feelings sometimes differs depending on the person. Therefore, it is considered to be difficult to communicate groove feelings to an amateur and they are considered to be a drum technique that is difficult to learn.

With this background, we state the two aims of this study. First, we define stochastic models for tight and loose groove feelings to give a quantitative definition

of groove feelings. Second, as an application of this model, we propose an evaluation system for the groove feelings evoked by drum sounds in which a monaural, polyphonic drum set recording is input and evaluation values regarding tight and loose are output. This system makes it possible to quantitatively evaluate a drum sound as tight or loose, which will help drummers to acquire groove feelings.

We define the models of groove feelings with reference to Okuhira *et al.*[3]. They analyzed tight and loose drum performances and calculated deviations of snare and bass drum hit-points from score-based hit-points for tight and loose and showed their averages and variances. They concluded that when drummers play a tight performance, they put the hit-points of snare and bass drums earlier than score-based hit-points and when they play a loose performance, they put them later. Using these parameters, we define stochastic models of tight and loose groove feelings.

As an application of these stochastic models, we propose an evaluation system for the groove feelings evoked by drum sounds. The deviations of snare and bass drums must be extracted from a monaural drum sound source since this system takes a monaural drum sound source as its input. Ohya *et al.*[4] extracted these deviations by non-negative matrix factorization (NMF)[5]. We refer to Ref.[4] to obtain deviations and propose our evaluation system by combining this method and our groove models.

The paper is organized as follows. Definitions of groove models for tight and loose and our proposed evaluation system are introduced in Section 2. The evaluation experiment and its results are described in Section 3, followed by a discussion. Finally, conclusions and directions for future work are presented in Section 4.

Table 1: Deviations as percentages of the length of a sixteenth note at a tempo of 100 for tight performance

	bass drum	snare drum
average	0.0641	0.150
standard deviation	0.0643	0.0358

Table 2: Deviations as percentages of the length of a sixteenth note at a tempo of 100 for loose performance

	bass drum	snare drum
average	-0.104	-0.157
standard deviation	0.0572	0.05

2 Method

2.1 Stochastic definitions of tight and loose groove feelings

One of the aims of this study is to define a stochastic model for both tight and loose groove feelings. Okuhira *et al.*[3] analyzed deviations of snare and bass drums hit-points from score-based hit-points. Tables 1 and 2 show percentages of deviations as percentages of the length of a sixteenth note at a tempo of 100 for tight and loose performances, respectively. These result suggest that these deviations are important features for distinguishing between tight and loose groove feelings.

On the basis of Okuhira *et al.*[4], we set these deviations as random variables of the models for tight and loose. We selected a bivariate normal distribution as the probabilistic distribution because it is considered that these deviations are distributed around the ideal deviation for both tight and loose owing to the fluctuation within a human performance.

If the average and variance of the deviation of snare and bass drums are (μ_s, σ_s^2) and (μ_b, σ_b^2) and the random variables are X_s and X_b , respectively, this model is defined as

$$\begin{pmatrix} X_s \\ X_b \end{pmatrix} \sim \text{MultiNormal} \left(\begin{pmatrix} \mu_s \\ \mu_b \end{pmatrix}, \begin{pmatrix} \sigma_s^2 & 0 \\ 0 & \sigma_b^2 \end{pmatrix} \right). \quad (1)$$

Since it can be considered that there is no correlation between deviations of snare and bass drums, the covariance is set to 0.

The parameters of these models are calculated using the data in Tables 1 and 2. These tables show deviations as percentages of the length of a sixteenth note at a tempo of 100. We assumed that these percentages are constant regardless of the tempo. Under this assumption, when we are given a tempo for a drum performance, the parameters of the model can be calculated from these percentages.

Fig. 1: Block diagram of our evaluation system

Fig. 2: Overview of matrices defined in NMF

2.2 Evaluation system for groove feelings

As an application of the tight and loose models, we propose a system for evaluating groove feelings. Figure 1 shows a block diagram of this system. Before applying NMF, Short-time Fourier transform (STFT) is performed to obtain an observation matrix. Then the snare and bass hit-points are calculated by NMF and then the deviations of these hit-points from score-based hit-points are calculated. Finally, we calculate the likelihood for tight and loose models and this can be considered the evaluation value of tight or loose groove feelings.

2.3 Non-negative matrix factorization

NMF[5] is an algorithm for decomposing a non-negative matrix into two non-negative matrices using its update algorithm. If a non-negative matrix $V \in R^{N \times M}$ is given, we initialize $W \in R^{N \times K}$ and $H \in R^{K \times M}$ with random numbers and update W and H to minimize a distance function, such as Euclidian distance or KL divergence.

Fig. 3: Score used in our evaluation experiment. At the top is shown hi-hat, snare, and bass drum hit-points as below

NMF has applications in many fields such as audio signal processing, image processing, astronomy, and bioinformatics. In particular, when it is applied to audio signal processing, $\mathbf{V}$ is called the observation matrix, which expresses the power spectrogram of an audio signal. Since a power spectrogram has non-negative values and can be thought of as a matrix, it can be decomposed by NMF. The decomposed $\mathbf{W}$ and $\mathbf{H}$ are called the observation matrix and basis matrix, respectively.

Figure 2 shows an overview of these matrices. Basis matrix $\mathbf{W}$ consists of column basis vectors that express the audio spectrum frequencies observed in the audio signal. Activation matrix $\mathbf{H}$ is considered to be a matrix which has row activation vectors that express the activation of the corresponding basis vector.

In this research, we assume that a drum performance consists of hi-hat, snare, and bass drums and set the value of K to three. Therefore, $\mathbf{W}$ is expected to have basis vectors that express hi-hat, snare, bass drum spectra, and $\mathbf{H}$ is expected to have row activation vectors that express activation of these instruments.

When the euclidean distance is used as the distance function, the update formula of NMF is written as

$$H_{km} \leftarrow H_{km} \frac{(W^T V)_{km}}{(W^T W H)_{km}}, \quad (2)$$

$$W_{nk} \leftarrow W_{nk} \frac{(V H^T)_{nk}}{(W H H^T)_{nk}}. \quad (3)$$

To obtain an accurate activation, $\mathbf{W}$ is fixed as a matrix $\mathbf{W}_0$, which has the hi-hat, snare, and bass drum spectra which is taken from the drum sound source of the Apple Garage Band. Therefore, we only update $\mathbf{H}$ as follows:

$$H_{km} \leftarrow H_{km} \frac{(W_0^T)_{km}}{(W_0^T W_0 H)_{km}}. \quad (4)$$

Note that the L_1 norm of the basis vectors in the basis matrix is normalized to one to avoid the scale optionality of the base.

The activation matrix is updated using Eq. (4), and activations of snare and hi-hat are obtained. These activations express the loudness of these instruments; thus, the local maxima are considered to be these hit-points.

Fig. 4: Activation matrix for tight performance

Fig. 5: Activation matrix for loose performance

2.4 Criterion for measuring deviation

To calculate the deviation of hit-points, a criterion for measuring the deviation is required. In this study, we assume that a performance has a fixed tempo, therefore, the time interval of the criterion is treated as a constant value and calculated from the tempo. The initial time of the criterion is calculated by subtracting the average deviation of the bass drum from the first hit-point of the bass drum for both tight and loose.

2.5 Evaluation of tight and loose

The log-likelihood function is defined as

$$\log L(\theta) = \log \prod_{i=1}^N f(x_i|\theta) = \sum_{i=1}^N \log f(x_i|\theta) \quad (5)$$

for the probability distribution $f(x|\theta)$ of the parameter θ in the case of N samples $x_1, \dots, x_N$.

Normally, parameter θ is handled as a variable in maximum likelihood estimation or other approaches. In this research, the parameters of this likelihood function are fixed for both tight and loose, and the obtained constant value is taken as the evaluation value for tight and loose.

Fig. 6: Peaks of snare and bass activation for tight performance

Fig. 7: Peaks of snare and bass activation for loose performance

3 Results

3.1 Experimental conditions

We conducted an evaluation experiment to determine this system correctly evaluates tight and loose performances. We instructed semi-professional drum player with more than ten years of drumming experience to play tight and loose performances while listening to a metronome at a tempo of 80. Conditions such as the tempo, the player and instruments used, and the score were all the same, and a polyphonic recording was made at a sampling rate of 192kHz. Figure 3 shows the score used in this experiment, which is considered to be the most basic score for a drum performance and displays hit-points of hi-hat, snare, and bass drums at the top.

3.2 Parameters of tight and loose models

The parameters described by Eq. (1) were calculated referring to Section 2.1. For the tempo of the recording of 80, the obtained parameters are shown in Tables 3 and 4.

Table 3: Model parameters of tight at tempo of 80 (s)

	snare	bass
average	0.00402	0.0121
standard deviation	0.0067	0.0121

Table 4: Model parameters of loose at tempo of 80 (s)

	snare	bass
average	-0.0295	-0.0194
standard deviation	0.0938	0.0107

3.3 Extraction of hit-points and calculation of deviation

First, STFT was performed on the sound source to obtain an observation matrix; then we performed NMF on this observation matrix and obtained the activation matrix. Figures 4 and 5 respectively show the activation matrix for tight and loose. Peak estimation was performed for snare and bass activation. Figures 6 and 7 show the results of snare and bass peak estimation. Setting an appropriate threshold, we obtained hit-points of snare and bass drums for tight and loose.

Referring to Section 2.5, we calculated the deviation of the snare and bass drum hit-points from actual hit-points for the performance at tempo of 80.

3.4 Evaluation to tight and loose

From the obtained deviations, we calculated log-likelihoods for tight and loose groove models and obtained evaluation values for tight and loose. Table 5 shows evaluation values of tight and loose drum performances.

3.5 Discussion

Table 5 shows that for the tight performance, the evaluation value of tight is larger than that of loose, and for the loose performance, the evaluation value of loose is larger than that of tight. Therefore, it is possible to identify tight and loose performances for the sound source used in this research. However, the many assumptions made regarding the drum performance must be generalized.

First, although we made a recording in which the drum player listened to a metronome, this situation is not realistic and we need to generalize this condition. We calculated the criterion from the metronome's tempo by assuming the performance tempo to be constant. However, the criterion must be calculated from the audio signal to generalize the assumption. By this generalization, we can consider increasing and decreasing of tempo and dynamics, such as accelerando, ritardando.

Second, we only treated three main types of drum, hit-hat, snare, and bass drums, that were played in the

Table 5: Evaluation of tight and loose

	tight performance	loose performance
tight evaluation	-2800	-509
loose evaluation	-3870	-170

most basic drum pattern. Because there are other instruments, such as tom-toms and cymbals, and many other scores for drumming, they must also be considered. Although we separated the drum audio signal by NMF, as described in Section 2.3, it is necessary to improve the method so that other instruments, and scores can be handled.

In addition to generalizing these assumptions, we must also evaluate the validity of the models. Groove feelings had not been defined quantitatively before our research, and as the first step, we defined the models for tight and loose as a bivariate normal distribution with deviations for snare and bass as random variables. However, the validity of these models has not been confirmed and should be evaluated in the future.

4 Conclusions

We defined tight and loose groove feelings and proposed an evaluation system for groove feelings evoked by drum sounds. The groove models were defined as a bivariate normal distribution with deviations for snare and bass drums as random variables. Using these models, we proposed our evaluation system of groove feelings, and as a result of the evaluation experiment, we showed that this system correctly identifies tight and loose groove feelings. Future work will be directed toward generalizing the assumptions and evaluating the validity of the models.

5 Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 17H00749 and JST PRESTO Grant Number JPMJPR1774, Japan.

References

- [1] C. Keil: “Motion and Feeling through Music,” *The Journal of Aesthetics and Art Criticism*, Vol. 24, No. 3, pp. 337-349 (1966).
- [2] S. Kawase, K. Eguchi: “The Concepts and Acoustical Characteristics of ‘Groove’ in Japan,” *PopScriptum 11 The Groove - Issue*, pp. 1-45 (2010).
- [3] K. Okuhira, K. Hirata, H. Katayori: “Relationship between Groove Feeling and the Timing and Loudness of Drum Attacks in Popular Music,” *IPSJ SIG Technical Report MUS*, pp. 21-27 (2004).
- [4] Y. Ohya, K. Nakamura, T. Tokunaga: “Extraction of Groove Feelings from Drum Data Using Non-negative Matrix Factorization,” *IEEE SCIS-ISIS*, pp. 125-130 (2012).
- [5] D. Lee, H. Seung: “Algorithms for Non-negative Matrix Factorization,” *NIPS* (2000).