A Digital Modeling Technique for Distortion Effect Based on a Machine Learning Approach

Yuto Matsunaga, Naofumi Aoki, Yoshinori Dobashi, and Tsuyoshi Yamamoto
Graduate School of Information Science and Technology, Hokkaido University, Sapporo, Japan
E-mail: matsunaga@ime.ist.hokudai.ac.jp

***Abstract*— This paper describes an experimental result of modeling stomp boxes of the distortion effect based on a machine learning approach. Our proposed technique models the distortion stomp boxes as a neural network consisting of CNN and LSTM. In this approach, CNN is employed for modeling the linear component that appears in the pre and post filters of the stomp boxes. On the other hand, LSTM is employed for modeling the nonlinear component that appears in the distortion process of the stomp boxes. All the parameters are estimated through the training process using the input and output signals of the distortion stomp boxes. The experimental result indicates that the proposed technique may have a certain potential to replicate the distortion stomp boxes appropriately by using the well-trained neural network.**

I. INTRODUCTION

Vintage stomp boxes still attract interest from many musicians because of their rare sound. It is difficult to obtain such sound without particular old electronic equipments that are no more manufactured nowadays. Therefore, as an application of digital signal processing technologies, digital modeling techniques have been studied in order to replicate the sound of such vintage stomp boxes.

For this purpose, many commercial products have been proposed [1,2]. However, the quality of such products is not always satisfactory. One of the reasons is that commercial products only simulate the representative characteristics of the reference stomp box in many cases, although each stomp box has unique characteristics due to its individual variation with different aging processes of the electronic circuits. To alleviate this problem, other techniques are still expected in order to improve the quality of the results simulated by the digital modeling techniques.

The digital modeling techniques may be categorized into two approaches. One is to simulate the real behavior of the electronic circuits of stomp boxes. Although this approach may result in physically collect output, it requires a lot of computational cost to replicate stomp boxes [3].

The other is to focus on the superficial characteristics of stomp boxes. Since this approach only models the relationship between the input and output signals of stomp boxes, it is called black box approach [4]. Compared with the electronic circuits simulation approach described above, this approach may be easier to perform, since it simplifies the entire process to be simulated. However this approach is not always appropriate especially if the process does not well represent the real behavior of the electronic circuits.

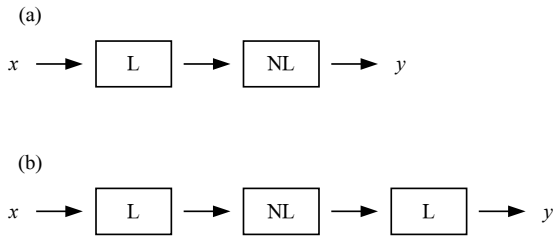

Fig. 1 Block diagrams for modeling distortion stomp boxes: (a) L-NL model, (b) L-NL-L model.

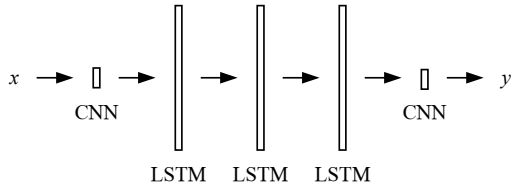

Fig. 2 Neural network representation of L-NL-L model

To mimic the characteristics of stomp boxes more effectively, the black box approach must cover various conditions that are not well taken into account in the conventional one. For this reason, we have investigated the possibilities of introducing a machine learning approach that estimates appropriate models from given data obtained from various conditions with ensuring the consistency of the estimated models.

II. CONVENTIONAL TECHNIQUE

As shown in Fig.1 (a), the conventional technique employs the block diagram that connects the linear and nonlinear components of the electronic circuits [4]. This L-NL model focuses on modeling the nonlinear component after removing the linear component included in its pre filter.

Although this approach gives acceptable results for modeling the nonlinear component itself, the conventional technique does not much consider the entire behavior of the distortion stomp boxes. As shown in Fig.1 (b), in general cases, the electronic circuits of the distortion stomp boxes include another linear component in its post filter. To give more realistic results, this L-NL-L model may potentially be an appropriate candidate for modeling the entire behavior of the distortion stomp boxes.

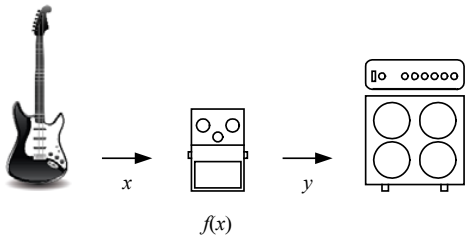

Fig. 3 Input and output signals of the distortion stomp box.

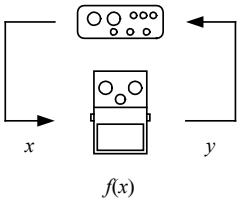

Fig. 4 Setup for the acquisition process of training data.

III. Proposed Technique

In this study, we have investigated the possibilities of introducing a machine learning approach for modeling the distortion stomp boxes.

As shown in Fig.2, the L-NL-L model may be represented as a multi layered neural network. The linear component may be modeled by a single node of CNN (Convolutional Neural Network) that represents a FIR (Finite Impulse Response) filter. The nonlinear component may be modeled by LSTM (Long Short Term Memory).

LSTM is a class of RNN (Recurrent Neural Network), a neural network that has recursive loops for learning the time series information [5]. Compared with the conventional RNN, LSTM is a better choice for learning the characteristics of time series information that have temporal contexts, since its structure appropriately treats the vanishment process about the past information that is not well taken into account in RNN.

LSTM is adopted in prediction problems that estimate future information from its past information [6]. It is employed for system identification problems that estimate the parameters of their assumed models.

As shown in Fig.3, the distortion stomp box is connected between a guitar and its amplifier. Its operation is mathematically expressed as $y = f(x)$, where x and y represent the input and output signals of the stomp box.

Our proposed technique models the stomp box as a neural network. It estimates $f(x)$ from the relationship between x and y through the training process of the neural network.

Figure 4 shows the setup for the acquisition process of training data. It is obtained from the input and output signals of the reference stomp box connected to an audio interface controlled by its host PC that sends the input signals and receives the output signals.

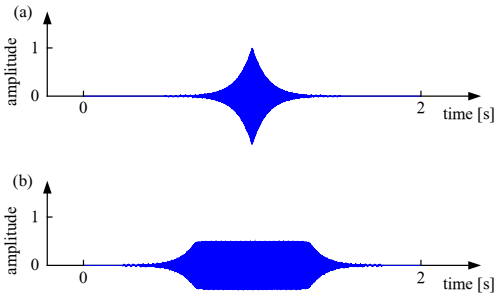

Fig. 5 Example of (a) input signal and (b) its corresponding output signal.

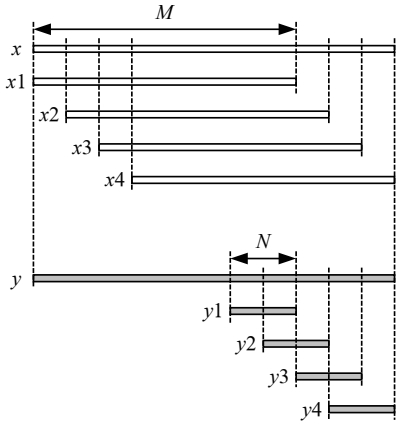

Fig. 6 Input and output data for the training process: $x1$ through $x4$ are input data, while $y1$ through $y4$ are their corresponding output data.

Since the characteristics of the distortion effect drastically changes according to the magnitude of its input level, the training data is obtained in the condition that sinusoids with varying amplitudes was employed as the input signals. It is defined as follows.

$$x(n) = a(n)\sin\left(\frac{2\pi f n}{f_s}\right) \quad (1)$$

where f_s is the sampling frequency, f is the frequency of the sinusoid covering the entire range of 59 electric guitar tones form E2 (82.4 Hz) to D7 (2349.3 Hz), and $a(n)$ is an envelope function that defines exponentially rising and falling amplitudes.

Figure 5 shows examples of the input and output signals. As shown in this figure, the output signal is obtained from the input signal by the clipping process of the distortion effect.

Figure 6 illustrates the relationship between the input and output data for the training process. Each input data consists of M samples obtained from the input signal of the stomp box. Its corresponding output data consists of N samples obtained from the output signal of the stomp box. All the training data are obtained by shifting the section of a pair of the input and output signals.

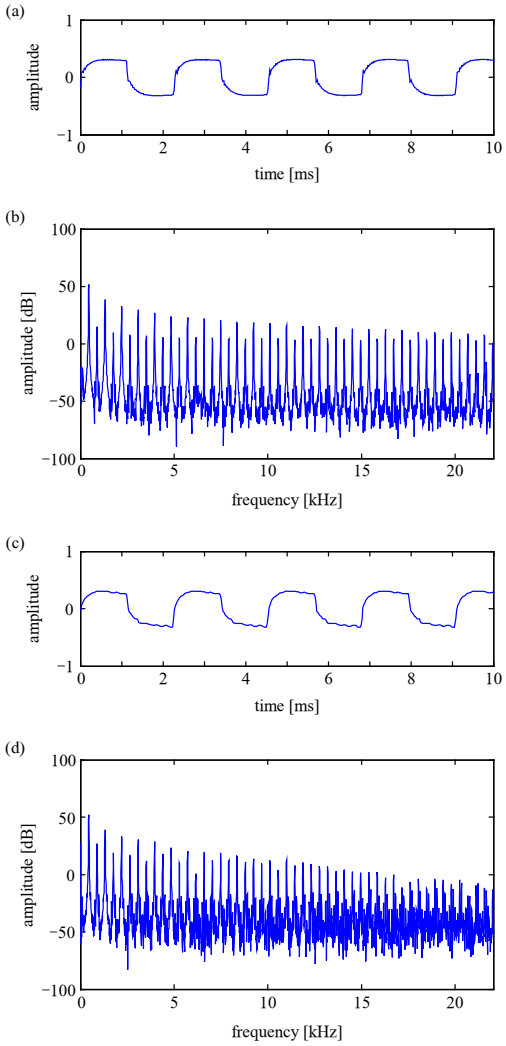

Fig. 7 Simulation with the input signal of 440 Hz sinusoid: (a) output signal of the reference stomp box, (b) its frequency characteristic, (c) output signal of the trained neural network, and (d) its frequency characteristic.

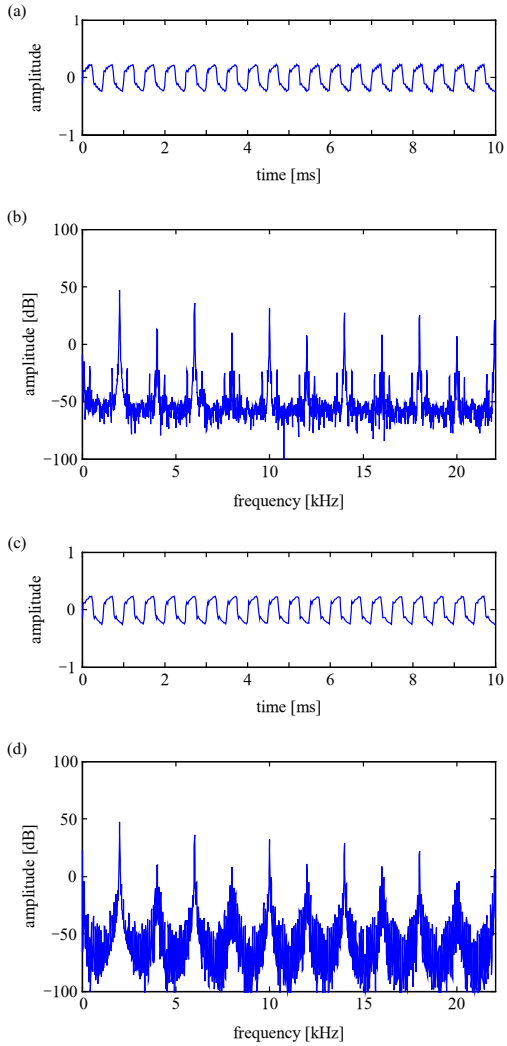

Fig. 8 Simulation with the input signal of 2000 Hz sinusoid: (a) output signal of the reference stomp box, (b) its frequency characteristic, (c) output signal of the trained neural network, and (d) its frequency characteristic.

IV. EXPERIMENT

To evaluate the proposed technique, as a pilot case, we applied the proposed technique to a commercial stomp box of the distortion effect (BOSS Blues Driver: BD-2). It is employed as the reference stomp box in the experiment.

For the data acquisition process, the sampling frequency and quantization level were set to be 44.1 kHz and 16 bits, respectively. Each knob of the reference stomp box was fixed at its average level during the data acquisition process.

The duration of the input and output signals was 2 seconds. The duration of the training data including the input and output signals in total was 118 s. For the training process, M and N were set to be 100 and 1, respectively.

The training process was implemented with TensorFlow with Keras running in Python. The order of each FIR filter modeled by CNN was set to be 1024. Three LSTM layer with 64 hidden neurons were adopted.

The epoch and batch size were set to be 10 and 300, respectively. The loss used in the training process was evaluated using MSE (Mean Squared Error) calculated from the output data of the neural network by comparing with the reference output data. It took about 9 hours for the training process with a GPU (Quadro K620). The final loss after the training process was 0.0043, while the initial loss before the training process 0.0289.

After the training process, simulations using arbitrary sinusoids were performed. Two results are shown in Fig.7 and Fig.8, respectively.

Figure 7 shows the result of the simulation with the input signal of 440 Hz sinusoid. This frequency was included in the training process. Compared with the output of the reference stomp box shown in this figure, the result indicates that the proposed technique appropriately mimics the waveform and frequency characteristics of the reference stomp box if the condition was included in the training process.

On the other hand, Figure 8 shows the result of the simulation with the input signal of 2000 Hz sinusoid. This frequency was not included in the training process. Compared with the output of the reference stomp box shown in this figure, the result indicates that the proposed technique also appropriately mimics the waveform and frequency characteristics of the reference stomp box even if the condition was not included in the training process.

Note that these results were obtained from an identical neural network. This indicates that a well-trained neural network may appropriately replicate the reference stomp box based on the machine learning approach.

To evaluate the proposed technique, an objective evaluation was conducted. Figure 9 shows the ESR (Error to Signal Ratio) between the reference output signal and the simulated output signal calculated in the frequency domain. The index is defined as follows.

$$ESR = \frac{\sum \left| X_{ref} - X_{sim} \right|^2}{\sum \left| X_{ref} \right|^2} \tag{2}$$

where X_{ref} represents the reference output signal, and X_{sim} the simulated output signal. In this figure, the horizontal represents the frequency of the input sinusoid, and the vertical corresponds to its ESR.

The conventional research indicates that the ESR of 0.1 or lower is considered to be acceptable quality [7]. In this sense, as shown in Fig.9, many cases were found to be acceptable, while there were a little exceptions.

To evaluate the proposed technique, another objective evaluation was also conducted. Figure 10 shows the correlation coefficient between the reference output signal and the simulated output signal calculated in the frequency domain. The index is defined as follows.

$$\rho = \frac{\mathrm{cov}\left(X_{ref}, X_{sim}\right)}{\mathrm{var}(X_{ref})\,\mathrm{var}(X_{sim})} \tag{3}$$

where X_{ref} represents the reference output signal, and X_{sim} the simulated output signals. In this figure, the horizontal represents the frequency of the input sinusoid, and the vertical corresponds to its correlation coefficient.

The conventional research indicates that the correlation coefficient of 0.95 or greater is considered to be acceptable quality [7]. In this sense, as shown in Fig.10, many cases were found to be acceptable, while there were a little exceptions.

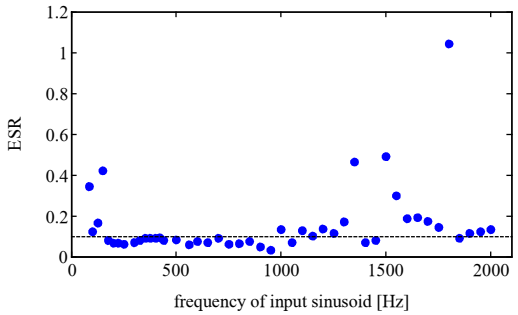

Fig. 9 ESR between the reference and simulated output signals. The dotted line shows the criterion in the conventional technique.

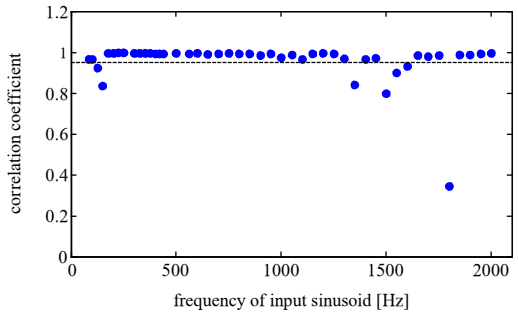

Fig. 10 Correlation coefficient between the reference and simulated output signals. The dotted line shows the criterion in the conventional technique.

V. DISCUSSION

The experimental result indicates that the proposed technique may appropriately generate output signals that are similar to those of the reference stomp box using a well-trained neural network.

However, the result shows that there are some deterioration especially in the high frequency region as shown in Fig.7 and Fig.8. This may be caused by the limitation of the sampling frequency in the data acquisition process. To improve the quality of the proposed technique, it is of interest to employ higher sampling frequency to decrease the influence found in the high frequency region.

Although the proposed technique is found to be almost acceptable, it shows that it is not always satisfactory for all the frequency employed in the evaluation. As shown in Fig.9 and Fig.10, there is the worst case with the input signal of 1800 Hz sinusoid in the pilot case. The reason of this problem may be attributable to the limitation of the training data. It should be clarified whether the results are improved if we prepare more data for the training process.

For more discussion, further experiments need to be carried out to evaluate effectiveness of the proposed technique. In addition, it seems that subjective evaluations also need to be carried out to confirm whether the proposed technique can surely mimic the timbre of the reference stomp box.

VI. Conclusions

In this paper, we propose a digital modeling technique for stomp boxes of the distortion effect based on a machine learning approach. The experimental result indicates that the proposed technique may have a certain potential to replicate the distortion stomp boxes appropriately by using the well-trained neural network.

The proposed technique does not assume any particular electronic circuits, so that it may be more flexible approach as a digital modeling technique, compared with the conventional techniques.

This paper only shows a pilot case. To improve the quality of the proposed technique, there still remain a number of agendas to be considered as future works.

One is to investigate why this approach yields the acceptable results. It is of interest to consider the relationship between the neural network model and the actual electric circuits of stomp boxes. It may give a clue to design the best structure of the neural network representation. The hyper parameters have not been tuned yet in the pilot case. Ablation study towards the best result is also necessary.

The other is to investigate how much data we need for the training process to obtain acceptable results. In the pilot case, we only employed the limited training data. It is of interest to confirm if the quality of the proposed technique is improved by increasing the number of the training data.

In addition, in the pilot case, the training process was performed with the single condition of the knobs setting of the reference stomp box. For the practical application, the training process should consider other conditions that take account of the changes of the knobs setting. It is also considered as one of the future works.

References

[1] Fractal Audio Systems, ABOUT, http://www.fractalaudio.com /about.php

[2] positivegrid, BIAS FX, https://www.positivegrid.com/biasfx/

[3] D.T. Yeh and J.O. Smith, “Automated physical modeling of nonlinear audio circuits for real-time audio effects: part 1 theoretical development,” IEEE Trans. Audio, Speech, and Language Process, pp. 203-206 (2010)

[4] F.Eichas and U.Zölzer, “Black-box modeling of distortion circuits with block-oriented models,” Digital Audio Effects (DAFx-16), pp.39-45 (2016)

[5] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, pp.1735-1780 (1997)

[6] J. Brownlee, “Long short-term memory networks with python develop sequence prediction models with deep learning,” Machine Learning Mastery (2017)

[7] F. Eichas, S. Moller, U. Zölzer, “Block-oriented gray box modeling of guitar amplifiers,” Digital Audio Effects (DAFx-17), pp.184 - pp 191 (2017)