IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 1 April 2023, accepted 20 May 2023, date of publication 29 May 2023, date of current version 6 June 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3280860

RESEARCH ARTICLE

Conditional Generative Adversarial Network-Based Tactile Stimulus Generation for Ultrasonic Tactile Display

KOKI HATORI[1] AND KENJIRO TAKEMURA[2], (Member, IEEE)
[1]Graduate School of Science for Open and Environmental Systems, Keio University, Yokohama 223-8522, Japan
[2]Department of Mechanical Engineering, Keio University, Yokohama 223-8522, Japan

Corresponding author: Kenjiro Takemura (takemura@mech.keio.ac.jp)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Bioethics Board of the Faculty of Science and Technology, Keio University.

ABSTRACT Tactile rendering is a promising technology that is necessary to integrate into virtual reality, augmented reality, mixed reality, and even metaverse environments. One of the key technologies for realizing tactile rendering is a reproduction or a display of tactile sensation. This study developed a model that generates an appropriate input signal to an ultrasonic tactile display using a conditional generative adversarial network. Sensory evaluation scores and vibration data acquired by a tactile sensor were used as training data for the conditional generative adversarial network-based models. In this study, different cluster analysis conditions were used to create the input information for the models. Each model generated the input signals for an ultrasonic tactile display, and the accuracy of the models was evaluated through sensory evaluation experiments. The results showed that model accuracy improved with moderate cluster classification and that the reproducibility of tactile sensation created with the models developed in this study was improved when compared with the reproducibility of tactile sensation created without the models.

INDEX TERMS Tactile display, machine learning, GAN, haptics, sensory reproduction.

I. INTRODUCTION

In recent years, technologies to reproduce highly realistic tactile sensations have been in demand in various fields such as virtual/augmented/mixed reality environments, e-commerce, entertainment, and teleoperated robot applications [1], [2]. Many researchers have been actively developing tactile reproduction or rendering technology for decades [3], [4], [5], [6], [7]. To reproduce tactile sensations, technologies to quantify the tactile sensation of an object and to display tactile sensations using a tactile display are necessary. To quantify tactile sensation, it has been suggested that focusing on vibration information at the time of object touch is an effective approach given the human tactile perception mechanism [8], [9], [10], [11]. Tactile displays using ultrasonic transducers have also been suggested [12], [13], [14] because of their compact structure and the potential for flexible predictability in vibration stimulation using amplitude modulation techniques [15]. Hence, it is important to focus on vibration as a sensing method for the quantification of tactile sensation, and an ultrasonic transducer is an effective way to apply vibration stimuli. However, it is inappropriate to input the sensed vibration directly to the ultrasonic transducer in order to reproduce tactile sensations. This is because the relationship between the quantified vibration information obtained by a tactile sensor and the corresponding input to an ultrasonic tactile display is not clear. Therefore, haptic rendering via tactile sensation, that is, the conversion from sensed vibration to tactile sensation and from tactile sensation to input to an ultrasonic tactile display, is needed. For this tactile rendering, previous studies have addressed the conversion from sensed vibration

The associate editor coordinating the review of this manuscript and approving it for publication was Zheng Chen.

to tactile sensation [8], [9], [10], [11]. This study considers the conversion from tactile sensation to display input, and focuses on conditional generative adversarial networks (CGANs) [16], [17], which are a type of generative adversarial network (GAN) [18], [19], as the conversion method. In fact, several studies applying GAN to tactile technology have been conducted in recent years [20], [21], [22], [23]. These studies indicate the effectiveness of using GANs to generate input signals for tactile displays. Although the use of human tactile sensation is unique in this study, its aim is still to generate input signals for an ultrasonic tactile display. Therefore, this study uses a GAN, specifically a CGAN, as the tactile rendering method. Using a CGAN instead of just a GAN enables us to generate signals that take human tactile sensations into account.

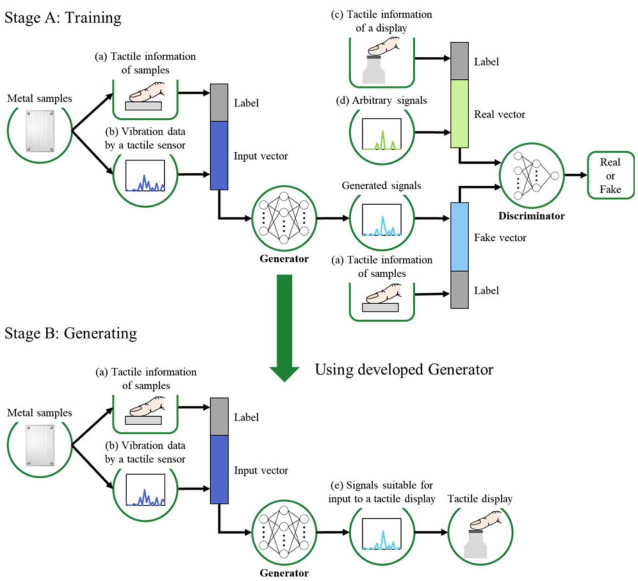

FIGURE 1. Concept of CGAN-based tactile stimulus generation model construction composed of Stages A and B for training and generating, respectively. (a) Tactile information of the metal samples is collected through sensory evaluation, (b) vibration data are measured by a tactile sensor by tracing it over the metal samples, and (c) tactile information when touching a tactile display generating (d) arbitrary signals is obtained by sensory evaluation. In Stage A, the CGAN-based tactile stimulus generation model is constructed using the datasets obtained through (a)–(d). In Stage B, (e) signals suitable for input to the tactile display are generated by the developed model.

II. CONCEPT OF CGAN-BASED TACTILE STIMULUS GENERATION MODEL

The concept of the CGAN-based tactile stimulus generation model is shown in Fig. 1. The model consists of two stages: training using the datasets (Stage A) and generating the input signals for an ultrasonic tactile display using the trained generator (Stage B). Metal plates are employed as test samples in this study. First, tactile information is quantified by a sensory evaluation with subjects by the semantic differential (SD) method (Fig. 1(a)). Simultaneously, vibration data are measured by a tactile sensor while it is tracing over a sample (Fig. 1(b)). In addition, as shown in Fig. 1(c), a sensory evaluation is conducted with arbitrary vibration stimuli (Fig. 1(d)) generated by the ultrasonic tactile display. By doing so, the relationship between the vibration stimuli generated by the ultrasonic tactile display and tactile evaluation score is obtained. The information obtained in the steps shown in Fig. 1(a)-(d) is used to create a dataset to construct a CGAN-based tactile stimulus generation model.

In the training part (Stage A), by alternately updating the parameters of the generator and the discriminator, the generator is eventually able to generate data corresponding to the conditional labels. In this study, sensory evaluation values were used for the conditional labels so that input signals to the ultrasonic tactile display can be generated based on human tactile information. Thus, by alternately training the generator and the discriminator, the generator learns what kind of input signals to generate for the tactile display in order to obtain the sensory evaluation values of the metal plate samples on the tactile display. The generated data approach the input signals that enable the tactile display to obtain sensory evaluation values similar to the values obtained when the samples are touched. Finally, in the generating part (Stage B), the constructed CGAN-based tactile stimulus generation model converts the quantified information into signals suitable for input to the ultrasonic transducer of the tactile display (Fig. 1(e)).

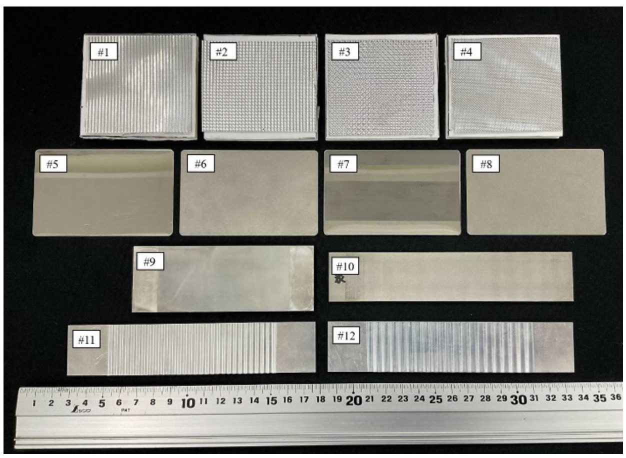

FIGURE 2. Twelve metal samples.

III. PREPARATION OF THE DATASETS

A. TACTILE EVALUATION OF METAL PLATE SAMPLES

A total of 32 people (10 females and 22 males) aged 22.5±0.9 years (range: 21-24 years) participated in the sensory evaluation experiment for the tactile sensations of 12 metal samples. The experiment protocol was approved in advance by the Bioethics Board of the Faculty of Science and Technology, Keio University. The participants were provided a thorough explanation of the evaluation methods and then signed an informed consent form before participating in the study. The metal samples are shown in Fig. 2. The subjects were asked to touch the samples one by one with visual information blocked, and then to score ten Japanese evaluation

TABLE 1. Words used in the sensory evaluation experiments words in brackets are in japanese.

Evaluation words (Japanese)		
Dry (Sarasara-suru)	Sleek (Subesube-suru)	Slippery (Tsurutsuru-suru)
Rugged (Gotsugotsu-suru)	Uneven (Bokoboko-suru)	Rough (Zarazara-suru)
Squishy (Gunyagunya-suru)	Prickle (Chikuchiku-suru)	Sticky (Petapeta-suru)
Rustle (Gasagasa-suru)		

words (shown in Table 1) on a 7-point unipolar scale ranging from 1 (negative) to 7 (positive) using the SD method. Visual information was blocked because human tactile evaluation is influenced by visual perception [24], [25]. Because the order of touch could influence the tactile evaluation, the evaluation order of the samples was random. The sensory evaluation experiment yielded 384 instances of 10-component one-dimensional data corresponding to the 10 evaluation words. The data are stored in IEEE DataPort (DOI: 10.21227/5jwn-qx81).

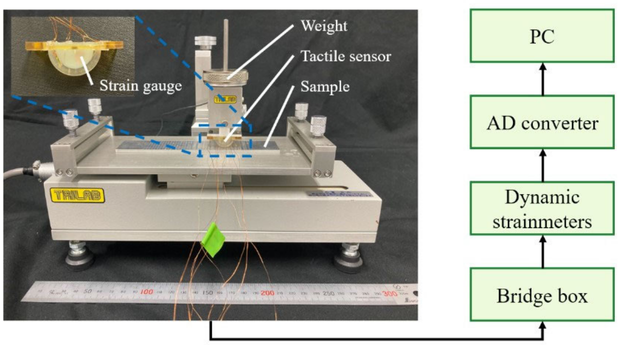

FIGURE 3. Overall view of the vibration data collection system.

B. VIBRATIONS WHEN SAMPLES ARE TRACED BY A TACTILE SENSOR

The vibration data collection system used in this study is shown in Fig. 3. A tactile sensor equipped with strain gauges inside it [10] is mounted on a friction tester (TL201s, Trinity-Lab Inc., Japan). The tactile sensor consists of a two-layer silicone rubber pad with different hardness, and two strain gauges are attached to phosphor bronze plates in each layer. Hence, there are a total of four strain gauges providing four vibration data from a single measurement. The outputs from the strain gauges are acquired by a PC. A weight ensures there is pressing force between the sensor and a sample while the friction tester moves the sample horizontally at a constant speed. The touch speed, measurement distance, vertical load, and sampling frequency were 20 mm/s, 50 mm, 0.98 N, and 10 kHz, respectively. Data were taken from the 2.5 s of measurement time during which the vibration waveform was stable for 2 s. Then, each vibration signal was standardized so that it had a mean of 0 and variance of 1. Twenty-one segments of 1-second data were cut from one length of standardized data by sliding a window 0.05 seconds per segment. Hence, 84 data were obtained per sample measurement because four types of vibration data were obtained in each measurement. Measurements were taken ten times for each of the 12 samples, and hence a total of 10,080 vibration data were obtained. Finally, FFT processing was performed on the vibration data obtained for each 1-s duration, which converted those data into an amplitude spectrum of 1,000-component one-dimensional data corresponding to frequencies ranging from 1 to 1,000 Hz. The data are stored in IEEE DataPort (DOI: 10.21227/5jwn-qx81).

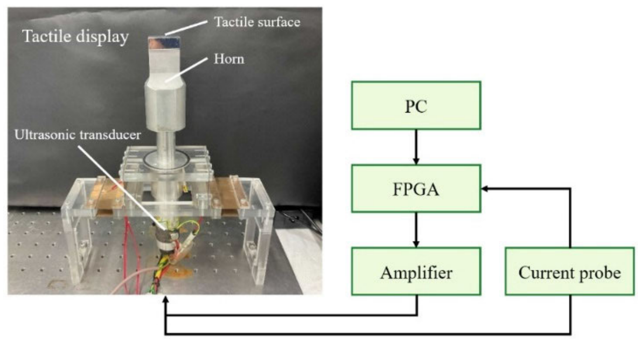

FIGURE 4. Overall view of the tactile display system.

C. TACTILE EVALUATION OF A TACTILE DISPLAY

An overall view of the tactile display system used in this study is shown in Fig. 4. In this study, a tactile display using a Langevin-type ultrasonic transducer was used. The ultrasonic transducer has a horn attached to its top to amplify the vibration amplitude. The tip of the horn is used as the touch surface. The touch surface is 15 mm long and 30 mm wide. In this study, tactile sensations were displayed by inputting amplitude-modulated waves into the ultrasonic transducer. The frequency of the carrier wave was fixed at the resonance frequency ($\approx$28.2 kHz) of the ultrasonic transducer. Various tactile sensations were displayed by changing the signal wave for amplitude modulated wave, which was generated using a field programmable gate array (FPGA; Analog Discovery 2, Digilent Inc., USA) for D/A conversion and controlled by PC software. The input waveform for the ultrasonic transducer was amplified by a gain factor of 10 using an amplifier (HAS4051, NF Corporation, Japan), resulting in a maximum voltage of 22 V being applied. To track the resonant state of the ultrasonic transducer even when it was being touched by a subject, the current flowing in the ultrasonic transducer was measured by the FPGA using a current probe (AC Current Probe CT2, Tektronix Inc., USA). The resonance tracking was conducted by adjusting the driving frequency until the admittance of the ultrasonic transducer was maximized.

Sensory evaluation experiments were conducted with a total of 32 people (10 females and 22 males) aged

22.5±0.9 years (range: 21-24 years) on the tactile display when arbitrary signals were input to the tactile display. The arbitrary signals are stored in IEEE DataPort (DOI: 10.21227/5jwn-qx81). The subjects were asked to touch the top surface of the tactile display and score ten Japanese evaluation words (shown in Table 1) on a 7-point unipolar scale from 1 (negative) to 7 (positive) using the SD method. Again, the order of signals input to the tactile display was random. The protocol was approved in advance by the Bioethics Board of the Faculty of Science and Technology, Keio University. The participants were provided a thorough explanation of the evaluation methods and then signed an informed consent form before participating in the study. The sensory evaluation experiment yielded 1248 instances of 10 component one-dimensional data corresponding to the 10 evaluation words. The data are stored in IEEE DataPort (DOI: 10.21227/5jwn-qx81).

IV. CGAN-BASED TACTILE STIMULUS GENERATION MODEL

A. STRUCTURE OF CGAN-BASED MODEL

First, for the generator, the input data were the vibration data after FFT processing, obtained as described in Section III-B, which are 1,000 component one-dimensional data corresponding to frequencies from 1 to 1,000 Hz. The conditional labels for the input data were the sensory evaluation values of the samples, obtained as described in Section III-A, which are 10 component one-dimensional data corresponding to the 10 evaluation words. Each sensory evaluation value used as a conditional label was normalized to a value between 0 and 1 (note that they were originally on a 7-point scale from 1 to 7). The above input data and conditional labels were combined, and one-dimensional data composed of 1,010 components were used as the input for the generator. The output is an amplitude spectrum of 1,000 component one-dimensional data corresponding to frequencies from 1 to 1,000 Hz generated by the generator. Table 2 shows the internal structure of the generator, where "Input" indicates the input layer, "Output" indicates the output layer, and the layers between the input and output layers are the intermediate layers. "Dense" represents a fully connected layer, and Leaky ReLU is a type of activation function, expressed as

$$f(u) = \begin{cases} au\,(u < 0) \\ au\,(u \geq 0)\,, \end{cases} \tag{1}$$

where u is the input, $f(u)$ is the output, and α is a constant coefficient ($\alpha = 0.01$ in this study). By employing a gradient to the negative ranges, a Leaky ReLU makes it possible to continue training neurons without stopping error backpropagation in the $u \leq 0$ range. Batch normalization [26] is a method that prevents gradient loss and divergence by standardizing the data distribution at each layer to mean 0 and dispersion 1. It can be used to stabilize learning and increase learning speed. Finally, dropout [27] is a method that prevents overlearning by inactivating neurons at a certain rate. In this study, the percentage of inactivation was set at 50%.

TABLE 2. Internal structure of the generator.

Generator	Number of neurons
Input: Vibration data+label	1,000+10
Dense	1,024
Leaky ReLU	1,024
Batch Normalization	1,024
Dense	512
Leaky ReLU	512
Batch Normalization	512
Dropout	512
Dense	256
Leaky ReLU	256
Batch Normalization	256
Dense	128
Leaky ReLU	128
Batch Normalization	128
Dropout	128
Dense	1,000
Output: Leaky ReLU	1,000

Next, for the discriminator, there are two types of input information. One is the 1,010-component one-dimensional data and combines the signals generated by the generator with the sensory evaluation values of the samples obtained as described in Section III-A as conditional labels. The other corresponds to the training data and consists of 1,010-component one-dimensional data that combines the arbitrary input signals to the ultrasonic transducer of the tactile display with the sensory evaluation values of the tactile display as the condition labels, as described in Section III-C. The output is a value between 0 and 1 and is the probability that the discriminator identified the input data as training data. Table 3 shows the internal structure of the discriminator. Finally, "Sigmoid" indicates the sigmoid function, which allows the output to be treated as a probability by adjusting the output data to a value between 0 and 1.

TABLE 3. Internal structure of the discriminator.

Discriminator	Number of neurons
Input: Vibration data+label	1,000+10
Dense	64
Leaky ReLU	64
Dense	32
Leaky ReLU	32
Dense	1
Output: Sigmoid	1

B. CLUSTER ANALYSIS OF THE DATASETS

Using the results of sensory evaluation experiments on metal plate samples, 12 different samples were classified before training the CGAN. Five patterns of classification were prepared, and the differences in model performance for each group were tested. Cluster analysis was used to classify the samples. Cluster analysis is a type of multivariate analysis that collects highly similar data from given datasets and classifies them. As a result of the cluster analysis, five groups were defined for the 12 metal plate samples: Group A with no clustering, Group B with two clusters, Group C with three clusters, Group D with four clusters, and Group E with all 12 samples classified into separate clusters.

TABLE 4. Hyperparameters.

Optimization Algorithm	Adam
	α=0.001
	β_1=0.9
	β_2=0.999
	ε=1e-8
Loss function	Binary cross entropy
Batch size	64
Iterations	200,000

C. MODEL TRAINING AND SIGNAL GENERATION

A separate CGAN was trained for each cluster in Groups A to E. The hyperparameters of the training are listed in Table 4. Adam was used for the optimization algorithm. The loss function was binary cross entropy, expressed as

$$E(t, y) = -tlogy - (1 - t)log(1 - t), \tag{2}$$

where y is the output value of the discriminator and t is the correct answer label. The discriminator's correct answer label can be 0 or 1, depending on the type of data input to the discriminator, but in either case, the loss can be accounted for.

As shown in Fig. 1, the generator is extracted after training for generating the desired input signals for the ultrasonic transducer of the tactile display. As shown in Table 4, the number of iterations was set in advance to 200,000, but the performance of the model is not always the best at the end of all iterations in CGAN training. For this reason, the number of iterations processed before the generator was extracted determined based on the loss function in training and the discriminator's correct response rate. Fig. 5 shows each example of the loss function and the discriminator's correct response rate during training. First, regarding the loss functions, the loss of the generator is relatively large and the loss of the discriminator is relatively small because the generator cannot generate data that incorporates the training data in the initial stage of training and the discriminator can easily distinguish between the generated and training data. As the training progresses, the generator can generate data that takes the training data into account, and the discriminator's correct response rate declines. The loss of the generator decreases and the loss of the discriminator increases. However, if the loss of the discriminator increases monotonically, it loses its function as a discriminator. Therefore, the loss of the generator must be small while the losses of both oscillate. Second, the discriminator's correct response rate shows the

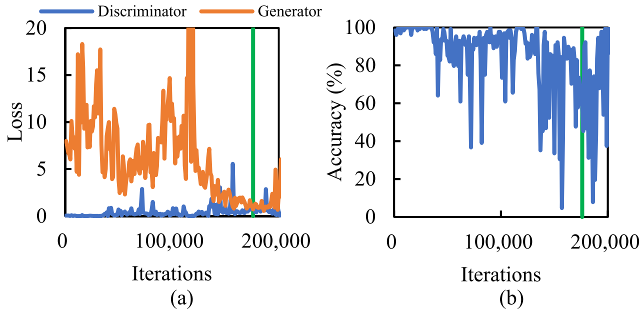

FIGURE 5. Examples of (a) the loss function and (b) the discriminator's correct response rate during training. These indicators were used to select the generator for data generation. The green line shows the number of iterations at which the generator was selected in this example.

percentage of data the discriminator correctly discriminated as generated data or training data in each iteration. In the early stages of training, the correct response rate is close to 100%, but as training progresses and the accuracy of the generator improves, the correct response rate declines. When the correct response rate is approximately 50%, the discriminator is unable to distinguish between the generated and training data. On the basis of the loss and discriminator's correct response rate, the generator was selected at a point in the iterations where the generator's loss was the smallest non-zero value and the discriminator's correct response rate was between 40% and 60%. The green line in Fig. 5 shows the iteration at which the generator was selected. If the above conditions were not satisfied during training, we defined the training as a failure. Groups A to E were used for training. Groups A to D were used to train the model successfully, whereas Group E, in which all 12 metal plate samples were classified separately, caused the training to fail.

For the models successfully trained using Groups A to D, the selected generator was used to generate signals for input to the ultrasonic transducer of the tactile display. Examples of input vibration data for the generator and the signal output from the generator are shown in Fig. 6. The output signals are the input signals to the ultrasonic tactile display.

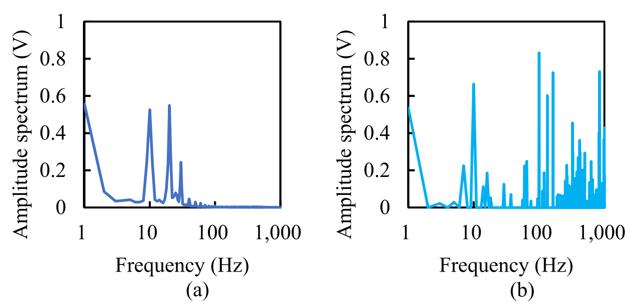

FIGURE 6. Each example of (a) input vibration data for the generator and (b) the signal output from the generator. The output signal is the input signal to the ultrasonic tactile display.

V. EVALUATION OF THE MODEL

A. EVALUATION EXPERIMENT

To evaluate the generated signals, sensory evaluation experiments were conducted to evaluate the tactile sensation when the signals generated by the models trained using Groups A to D were input to the ultrasonic transducer of the tactile display. No evaluation experiment was conducted with the Group E model because the training was not successful. In contrast, sensory evaluation experiments were also conducted for the tactile sensation when the vibration information acquired from the tactile sensor was directly input to the ultrasonic transducer. Twenty adult subjects (7 females and 13 males) aged 22.4±1.0 years (range: 21-25 years) were asked to evaluate the 10 evaluation words shown in Table 1 on a 7-point unipolar scale ranging from 1 (negative) to 7 (positive) using the SD method. Because the order of touch influences the tactile evaluation, the order of the signal input was random. The protocol was approved in advance by the Bioethics Board of the Faculty of Science and Technology, Keio University. The participants were provided a thorough explanation of the evaluation methods and then signed an informed consent form before participating in the study.

After the experiment, to confirm the change in tactile reproducibility with respect to the models, we calculated the mean absolute error between the tactile evaluation points with the generated signals and those obtained from the sensory evaluation of the samples. The results for each mean absolute error are shown in Fig. 7. The first column in Fig. 7 shows the mean absolute error between the tactile evaluation points when the vibration information acquired by the tactile sensor was directly input to the ultrasonic transducer of the tactile display and the tactile evaluation points of the samples.

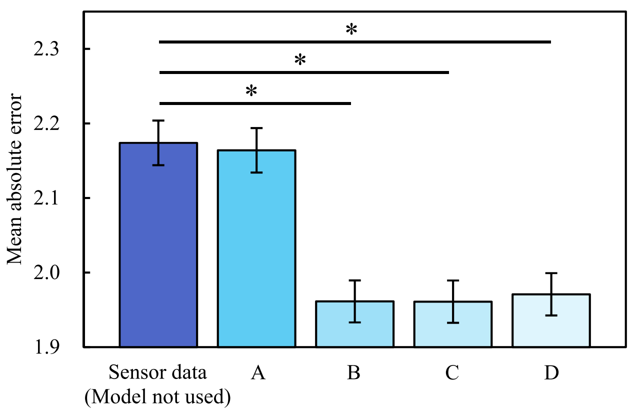

FIGURE 7. Mean absolute error from the tactile evaluation points of the metal plate samples. Statistical analysis showed that models trained using Groups B, C, and D had significantly smaller errors than the sensor data when the model was not used ($n = 2{,}400$, mean ±SE, *: $p < 0.01$).

The remaining columns in Fig. 7 correspond to the mean absolute errors between the tactile evaluation points when the signals generated by the models trained by Groups A to D were input to the ultrasonic transducer of the tactile display and the tactile evaluation points of the samples. Smaller errors indicate that the tactile evaluation is closer to the actual tactile sensation of the samples. To verify the statistical differences, Student's t-tests were conducted on the errors for each column. As a result, no significant difference in error reduction was observed for the model preprocessed using cluster analysis. However, the models trained using Groups B, C, and D, in which the samples were classified by cluster analysis, showed significantly reduced errors at the 1% level of significance. Therefore, for the Groups B, C, and D models, their use improved the reproducibility of tactile sensation, suggesting the effectiveness of the model constructed in this study. The model trained using Group A, which did not use cluster analysis, showed no significant difference in error when compared with the model that did not use the CGAN-based model, suggesting that cluster classification improved the performance of the CGAN-based tactile stimulus generation model.

VI. DISCUSSION

Cluster classification improved the performance of the model. It allows the models to generate signals that take into account the characteristics of each cluster, which is thought to improve the reproducibility of tactile sensations. First, for Group A, for which model training and signal generation were conducted without cluster analysis, a single model generated signals corresponding to the 12 samples. The signals were generated with small errors for all metal plate samples, and similar signals were output for all 12 types of samples. In contrast, for Groups B to D, which were used for training and generating signals after cluster analysis, models were constructed for each cluster and generated signals with smaller errors for each cluster. Thus, the use of cluster classification improved the performance of the model. However, we do not infer that a more detailed classification by cluster analysis leads to a smaller error. This is because there is no significant difference in the errors for Groups B, C, and D, which are classified into two, three, and four clusters, respectively. Furthermore, Group E, in which all samples were used to train a separate model, failed at the training stage. The reason for the lack of performance improvement with increasingly detailed classification is thought to be the reduction in the training datasets due to classification. The above suggests the use of cluster classification before training the model is effective, but it is necessary to keep the classification moderate.

VII. CONCLUSION

In this study, based on the tactile sensor and human tactile evaluation, we constructed CGAN-based tactile stimulus generation models to generate input signals for the ultrasonic transducer of the tactile display. First, to generate datasets for training, we conducted experiments to measure the vibration data of metal plate samples when they were touched by the tactile sensor. Sensory evaluations of the tactile sensation of the tactile display using an ultrasonic transducer and the metal plate samples were also conducted. Next, cluster analysis was performed based on the results of sensory evaluation experiments on metal plate samples, defining five groups with different degrees of classification. CGAN training was performed for each of the five defined groups to generate the input signals for the ultrasonic transducer of the tactile display, using the loss function of the training and the discriminator's correct response rate as indicators. Note that, with the finely classified clusters, i.e., each sample is used to train a separate model, the training did not succeed and signals could not be generated. Finally, an experiment was conducted to evaluate the signals generated by each model. The results showed that the reproducibility of tactile sensation generated with the CGAN-based tactile stimulus generation models with cluster classification was improved when compared with the reproducibility of tactile sensation generated without the model.

REFERENCES

[1] M. Sreelakshmi and T. D. Subash, "Haptic technology: A comprehensive review on its applications and future prospects," *Mater. Today, Proc.*, vol. 4, no. 2, pp. 4182–4187, 2017.

[2] D. Wang, Y. Guo, S. Liu, Y. Zhang, W. Xu, and J. Xiao, "Haptic display for virtual reality: Progress and challenges," *Virtual Reality, Intell. Hardw.*, vol. 1, no. 2, pp. 136–162, Apr. 2019.

[3] D. G. Caldwell, N. Tsagarakis, and C. Giesler, "An integrated tactile/shear feedback array for stimulation of finger mechanoreceptor," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 1999, pp. 287–292.

[4] D. Allerkamp, G. Böttcher, F.-E. Wolter, A. C. Brady, J. Qu, and I. R. Summers, "A vibrotactile approach to tactile rendering," *Vis. Comput.*, vol. 23, no. 2, pp. 97–108, Jan. 2007.

[5] Y. Asai, Y. Yokokura, and K. Ohishi, "Fine force reproduction of environmental haptic sensations based on momentum control," *IEEE Trans. Ind. Electron.*, vol. 63, no. 7, pp. 4304–4313, Jul. 2016.

[6] S. Bochereau, S. Sinclair, and V. Hayward, "Perceptual constancy in the reproduction of virtual tactile textures with surface displays," *ACM Trans. Appl. Perception*, vol. 15, no. 2, pp. 1–12, Feb. 2018.

[7] D. Degraen, M. Piovarči, B. Bickel, and A. Krüger, "Capturing tactile properties of real surfaces for haptic reproduction," in *Proc. 34th Annu. ACM Symp. User Interface Softw. Technol.*, Oct. 2021, pp. 954–971.

[8] E. Asaga, K. Takemura, T. Maeno, A. Ban, and M. Toriumi, "Tactile evaluation based on human tactile perception mechanism," *Sens. Actuators. A, Phys.*, vol. 203, pp. 69–75, Dec. 2013.

[9] L. Nobuyama, Y. Kurashina, K. Kawauchi, K. Matsui, and K. Takemura, "Tactile estimation of molded plastic plates based on the estimated impulse responses of mechanoreceptive units," *Sensors*, vol. 18, no. 5, p. 1588, May 2018.

[10] F. Ito and K. Takemura, "A model for estimating tactile sensation by machine learning based on vibration information obtained while touching an object," *Sensors*, vol. 21, no. 23, p. 7772, Nov. 2021.

[11] M. Sagara, L. Nobuyama, and K. Takemura, "Nonlinear tactile estimation model based on perceptibility of mechanoreceptors improves quantitative tactile sensing," *Sensors*, vol. 22, no. 17, p. 6697, Sep. 2022.

[12] F. Giraud, M. Amberg, B. Lemaire-Semail, and C. Giraud-Audine, "Using an ultrasonic transducer to produce tactile rendering on a touchscreen," in *Proc. Joint IEEE Int. Symp. Appl. Ferroelectr., Int. Workshop Acoust. Transduction Mater. Devices Workshop Piezoresponse Force Microsc.*, May 2014, pp. 1–4.

[13] D. A. Torres, B. Lemaire-Semail, C. Giraud-Audine, F. Giraud, and M. Amberg, "Design and control of an ultrasonic surface haptic device for longitudinal and transverse mode comparison," *Sens. Actuators A, Phys.*, vol. 331, Nov. 2021, Art. no. 113019.

[14] C. Sun, W. Nai, and X. Sun, "Tactile sensitivity in ultrasonic haptics: Do different parts of hand and different rendering methods have an impact on perceptual threshold?" *Virtual Reality, Intell. Hardw.*, vol. 1, no. 3, pp. 265–275, Jun. 2019.

[15] T. Watanabe and S. Fukui, "A method for controlling tactile sensation of surface roughness using ultrasonic vibration," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 1995, pp. 1134–1139.

[16] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.

[17] G. Antipov, M. Baccouche, and J. Dugelay, "Face aging with conditional generative adversarial networks," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2089–2093.

[18] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020.

[19] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 53–65, Jan. 2018.

[20] Y. Ujitoko, Y. Ban, and K. Hirota, "GAN-based fine-tuning of vibrotactile signals to render material surfaces," *IEEE Access*, vol. 8, pp. 16656–16661, 2020.

[21] S. Cai, L. Zhao, Y. Ban, T. Narumi, Y. Liu, and K. Zhu, "GAN-based image-to-friction generation for tactile simulation of fabric material," *Comput., Graph.*, vol. 102, pp. 460–473, Feb. 2022.

[22] C. Hernandez-Mejia, X. Ren, A. Thabuis, J. Chavanne, P. Germano, and Y. Perriard, "Generative adversarial networks for localized vibrotactile feedback in haptic surfaces," in *Proc. 24th Int. Conf. Electr. Mach. Syst. (ICEMS)*, Oct. 2021, pp. 105–110.

[23] Y. Ban and Y. Ujitoko, "TactGAN: Vibrotactile designing driven by GAN-based automatic generation," in *Proc. SIGGRAPH Asia Emerg. Technol.*, Dec. 2018, pp. 1–2.

[24] I. Motoyoshi, S. Nishida, L. Sharan, and E. H. Adelson, "Image statistics and the perception of surface qualities," *Nature*, vol. 447, no. 7141, pp. 206–209, Apr. 2007.

[25] Y. Ujitoko and Y. Ban, "Survey of pseudo-haptics: Haptic feedback design and application proposals," *IEEE Trans. Haptics*, vol. 14, no. 4, pp. 699–711, Oct./Dec. 2021.

[26] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, Jul. 2015, pp. 448–456.

[27] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, Jan. 2014.

KOKI HATORI received the B.S. degree in mechanical engineering from Keio University, Yokohama, Japan, in 2022, where he is currently pursuing the M.S. degree in mechanical engineering with the Graduate School of Science for Open and Environmental Systems.

His current research interests include haptics, tactile sensor, tactile display, and machine learning, in particular, generative adversarial networks.

KENJIRO TAKEMURA (Member, IEEE) received the B.E. degree in mechanical engineering, the M.E. degree in biomedical engineering, and the Ph.D. degree in integrated design engineering from Keio University, Yokohama, Japan, in 1998, 2000, and 2002, respectively. He joined the Department of Mechanical Engineering, Keio University, in 2002, as a Research Associate. From 2003 to 2008, he was an Assistant Professor with the Precision and Intelligence Laboratory, Tokyo Institute of Technology. Since 2008, he has been with the Department of Mechanical Engineering, Keio University, where he is currently a Professor. His research interests include soft robots, haptic interfaces, and biomedical devices. He received the JSME Young Engineers Award for research on multi-DOF ultrasonic motors, in 2002, the Best Research Paper Award from the 10th International Conference on Mechatronics Technology, in 2006, and the Best Paper Award from the ASME International Mechanical Engineering Congress and Exposition, in 2015, among others.