Automated Vibrotactile Generation based on Texture Images or Material Attributes using GAN

Yuki Ban\textsuperscript{1} Yusuke Ujitoko\textsuperscript{2,3}

\textsuperscript{1} Graduate School of Frontier Sciences, The University of Tokyo
\textsuperscript{2} Hitachi, Ltd.
\textsuperscript{3} Graduate School of Informatics and Engineering, The University of Electro-Communications

Keywords: Haptic display, Vibration design,

ABSTRACT

We propose the vibrotactile feedback designing system using GAN (Generative Adversarial Network)-based vibrotactile signal generator (TactGAN). Our system generate signals presenting specific tactile impression based on user-defined parameters. It can also automatically generate signals presenting the tactile impression of images. It realizes the rapid designing of vibrotactile signals for application with such feedback. User studies showed that it was not possible to distinguish between vibrations generated using this model and vibrations recorded from the actual material surface, and that the generated vibrations provided a sense of reality that was comparable to the actual vibrations.

1 INTRODUCTION

Recently, touch devices with vibrotactile feedback (VF) have been developed. VF is effective when it gives the tactile impression of the specific user’s operations (e.g. presenting hardness during pushing a button), and when it conveys the tactile impression of the specific contents in applications (e.g. presenting roughness of a cloth displayed on screen). However, most of the current applications on such devices don’t use VE effectively.

The main reason for this is that it takes much time to prepare appropriate vibrotactile signals for applications. To prepare the signals conventionally, application developers search the signals from a database that is a collection of recorded signals [1]. However, it is rare that developers could find the appropriate signals that meet requirements because the signal space is too large. When they cannot find the signals in the database, they must record them using actual surfaces. For this purpose, it is necessary to repeat a series of tasks such as preparing multiple materials to be recorded, recording the vibrations when tracing or tapping the materials using a vibration recording tool [2], evaluating the texture when the recorded vibrations are presented, and if they are not the desired sensation, testing with another material. It is the time-consuming task for developers to record every needed signal.

To solve these problems, we focus on the Generative Adversarial Networks (GANs) [3]. It has shown promising results in synthesizing real-world images. Prior research demonstrated that GANs could generate images conditioned on labels and texts [4].

The GAN framework should be able to generate desired vibrations with attributes and images as conditions. In this case, it can be said that there is a possibility to generate not only the attribute and the image used for the learning but also the targeted vibration when the attribute and the image synthesized from those attributes and the image are inputted. However, in spite of these previous research, there are few attempts to generate time-series data using GAN. Especially, generation of vibration by GAN has not been realized until now. As a framework, GAN is known to be unstable and difficult to learn and a lot of learning techniques has been developed to mitigate it. Since the data structure which has been the learning object in GAN until now mainly depends on the two-dimensional structure such as the image, it cannot be directly utilized for the learning of the time series data which is the one-dimensional data. This is one of the reasons why vibration generation by GAN has not been realized until now.

Thus, we proposed the vibrotactile designing system using a GAN-based vibrotactile generator (TactGAN) (Fig. 1). We made full use of GANs for indirectly generating vibrotactile signals via time-frequency domain representation, which can be calculated as the two dimensional image. Our TactGAN returns the appropriate signals directly generated from user-defined tactile impressions such as softness or roughness. Also, TactGAN returns the signals by user input images. The use of TactGAN would reduce the time taken to prepare the signals, and promote the development of...
applications with VF.

2 CONSTRUCTION OF MODEL FOR GENERATING VIBROTACTILE STIMULI

2.1 Overview

The overall diagram of our model is shown in Fig. 1. It consists of two parts: an encoder network, and a generator network. They are trained separately. The encoder is trained as an image classifier and it encodes texture images into a label vector \( c \). The generator is trained with discriminator in GANs training framework and generates spectrogram that is a representation of vibration in a time-frequency domain. The model at the time of learning consists of two components, the Generator and the Discriminator. The Generator and the Discriminator are learned simultaneously in the GAN learning framework. The details of learning will be described in the next section. Contrary to the learning phase, the Discriminator is removed and only the Generator is used in the inference phase. We describe the training details for each network in the following sections. The overall model enables end-to-end generation from visual images or label attributes of texture to the vibrotactile wave.

The input into the model is either a class label that represents the tactile attributes of the texture or a texture image. When the image is input, the label vector \( c \) is extracted from the texture image through encoder network. The label vector \( c \) is a categorical variable that shows the attributes of the texture. Next, the label vector \( c \) is passed into the generator network. The generator concatenates the label \( c \) and the random noise \( z \) and transforms them into the spectrogram. The generated spectrogram is converted into the acceleration wave format by Griffin-Lim algorithm [5]. Then the wave format data is output to the user. With this overall model, users can input either label information or texture images to obtain vibration.

Acceleration signals are used as vibrotactile stimulus in our model. In order to train the whole network, we use dataset [1], which contains acceleration signals and captured images during movement task. The pairs of signals and images are annotated with 108 classes.

2.2 Encoder

We trained the image encoder that encoded texture images into the label vector \( c \). We adopted the deep residual network (ResNet-50) [6] architecture. We fine-tuned all the layers of the ResNet-50 that had been pre-trained with ImageNet [7]. We used Adam optimizer with a mini-batch size of 64. The learning rate started from 1e-3 and was decreased by a factor of 0.1 when the training error plateaued.

The size of provided images by [1] is 320 x 480. We fed them into the encoder network. For training phase of encoder network, we followed ordinary data augmentation settings. We scaled an image with factors in [7], randomly cropped 128 x 128 size of it, flipped it horizontally and vertically, rotated it by a random angle. The recent data augmentation technique of random erasing and mixup were also used.

As a result of training, the trained encoder achieved a classification accuracy of more than 95 percent on the testing set. After the network was trained, its last layer was removed, and the feature vector of the second to the last layer having dimension of label vector was used as the image encoding in our generator network.

2.3 Generator

Generator was trained with discriminator in GANs framework. During training, the discriminator learned to discriminate between genuine and generated samples, while the generator learned to fool the discriminator. Generator output samples \( x = G(z, c) \) conditioned on both random noise vector \( z \) and a label vector \( c \) from dataset. Discriminator had two outputs: \( D(x) \) the probability of the sample \( x \) being genuine, and \( P(x) = c \), the predicted label vector of \( x \). After training, the discriminator was removed and the generator was only used in our model. Inspired by [8], we employed architecture and loss function, which was based on SRRResNet, DRAGAN, and AC-GAN. The architecture of generator and discriminator are shown in Fig. 2.

![Fig. 2 Network architecture.](image)

![Fig. 3 Selected textures for GANs training.](image)

![Fig. 4 The spectrogram for each class label in test dataset and the one generated by generator.](image)
Fig. 5 Experimental Settings

Acceleration signals orthogonal to the surface during movement task were used as vibrotactile stimulus and we aim at generating the signals by generator. For now, there are few studies generating time series data using GANs. It is because GANs are poor at generating time-series data though they are good at generating 2D images. Therefore, we chose amplitude spectrogram as a representation of the acceleration signals and trained GANs to generate spectrogram as if that was 2D image. The same dataset used for training encoder contained acceleration signals during movement task. Each signal had 4 seconds long and the sampling rate was 10 kHz. We computed the spectrogram from wave format using 512-point Short-Time Fourier Transform (STFT) with a 512 hamming window and a 128 hop size. Then, the linear amplitude of the spectrogram was converted to the logarithmic scale. We cropped the spectrogram and resized it into 128 x 128 size. As a result, the spectrogram contained the information of time-frequency domain up to 256 Hz for 1.625 seconds long.

We selected 9 textures out of 108 textures for GANs’ training because it is stable to train conditional GANs with fewer number of conditional label dimensions. Thus, the dimension of categorical label c was 9. On the other hand, the dimension of noise z was 50. The selected 9 textures were representative of 9 groups of LMT haptic texture database [1] (Fig. 3). We used Adam optimizer with a minibatch size of 64. The learning rate was fixed at 2e-4. The spectrogram in test dataset and the one generated by generator are shown in Fig. 4.

3 USER STUDY

User studies were conducted to investigate that whether our method could generate perceptually realistic vibrotactile stimuli. Two studies were conducted to evaluate Generator (Generator Ex.) and E2E network (E2E Ex.). Ten participants whose ages ranged from 22 to 25 (eight males and two females) participated in these studies.

3.1 Experimental System

In user studies, participants’ task was to move a pen-type device on a surface of a tablet device while receiving vibrotactile feedback. Our experimental system was constituted of the tablet device (Apple Inc., iPad Pro 9.7 inch), an amplifier (Lepai Inc., LP-2020A +), and a pen-type device with a vibrator (Fig. 5). The pen-type device, which we handcrafted, is specifically described in the next paragraph. The pen-type device was about 20 g weight and about 140 mm long. The pen tip wore conductive material that is ordinary used for the stylus. Since the shaft of the pen used in these studies was made of plastic and does not conduct to the grip part, we wired a conductive sheet on the grip to react with a capacitance type touch screen. Inside the pen-type device, the vibrator (ALPS Inc., HAPTIC Reactor) was embedded at the position of 2cm distance from the tip of the pen where participants gripped.

When participants touched and moved the pen on the surface, the vibration signal was output from earphone jack of the tablet, and amplified by the amplifier, and vibrator embedded on the pen presented the vibration to the participants’ fingers.

3.2 Task Design

These studies used a within-participant design. Participants moved the pen-type device along the two different predefined path on screen in succession, while receiving either test or generated vibrational feedback (Fig. 6). After that, participants tried to distinguish which stimulus was generated one. They also evaluated the realism of stimuli. In Generator Ex., generated signals were generated by feeding a label vector that represented each class into the generator. In E2E Ex., generated signals were generated by feeding a test image that represented each class into the encoder network. Corresponding class label texts or texture images were displayed on the touch screen. Participants’ task was the same in Generator Ex. and E2E Ex. except that what they saw on screen was class label texts or texture images.

The procedure of one trial in participant’s task is described in this paragraph. Participants moved the pen on a virtual texture surface from left to right for about 100 mm distance at fixed speed with their dominant hands. To control the movement speed and distance, the touch screen visualized a bar that indicated where and how much speed to move. According to the bar elongation, participants moved the pen approximately 100 mm distance in 1.6 seconds. Participants were told to hold the pen at the position where a vibrator was embedded. After completing movement on two surfaces, they answered which stimulus was felt generated one by
participants tended to misunderstand the generated stimulus as an genuine stimulus for these two classes. This results suggests that it is difficult to present the feeling of tracing a hairy or soft material with our system.

4 CONCLUSIONS
In this study, we constructed a model which outputs vibration using material attributes as input using GAN. Using this model, vibration can be collected even when users don’t have material itself or when the desired material does not vibrate in the data set. Through user study, it was shown that the vibration generated by the model cannot be distinguished from the true vibration under almost all conditions, and that the generated vibration is perceived to be as realistic as the true vibration. In the future, we will evaluate the generated vibration in more detail when input data are interpolation vectors of multiple materials that are not used for learning.

REFERENCES