2025 IEEE World Haptics Conference (WHC)
Suwon Convention Center, Suwon, Korea
July 8 ~ 11, 2025

Multi-Channel Vibrotactile Signal Compression Based on Implicit Neural Representation

Yuto Ozawa
School of Engineering
The University of Osaka
Suita, Osaka, Japan
ozawa.yuuto@ist.osaka-u.ac.jp

Akihiro Kuwabara
Grad. Sch. of IST
The University of Osaka
Suita, Osaka, Japan
kuwabara.akihiro@ist.osaka-u.ac.jp

Tom Ogura
Grad. Sch. of IST
The University of Osaka
Suita, Osaka, Japan
ogura.tom@ist.osaka-u.ac.jp

Takuya Fujihashi
Grad. Sch. of IST
The University of Osaka
Suita, Osaka, Japan
tfuji@ist.osaka-u.ac.jp

Shunsuke Saruwatari
Grad. Sch. of IST
The University of Osaka
Suita, Osaka, Japan
saru@ist.osaka-u.ac.jp

Takashi Watanabe
Grad. Sch. of IST
The University of Osaka
Suita, Osaka, Japan
watanabe@ist.osaka-u.ac.jp

***Abstract*—Haptic feedback is critical for enhancing immersion in extended reality applications. To enable realistic and immersive experiences, a compact representation of multi-channel vibrotactile signals is essential for storage and streaming. To obtain the compact representation, existing vibrotactile coding schemes leverage psychohaptic models to reduce perceptual redundancy in the haptic domain. However, the coding efficiency is still low. In this paper, we integrate the psychohaptic model and the implicit neural representation (INR), i.e., the power of neural networks (NNs), to obtain a further efficient representation for multi-channel vibrotactile signals. For this purpose, the proposed scheme overfits the multi-channel vibrotactile signals to a small NN and regards the overfitted weights of the NN as the representation of the vibrotactile signals. In addition, we propose a psychohaptic-inspired loss function for training the proposed NN architecture to obtain the compact representation with a slight degradation of the user's perceptual quality. Experiments using an open multi-channel vibrotactile dataset and the user perception-aware metric demonstrate that the proposed scheme simultaneously achieves compact and less quality-distorted vibrotactile representations compared with the state-of-the-art vibrotactile coding schemes.**

***Index Terms*—vibrotactile signals, implicit neural representation, psychohaptic model**

I. Introduction

Haptic feedback [1] is a key modality to improve immersion through extended reality applications by introducing the feedback in addition to visual and audio stimuli. Many studies have exploited an arrayed (multi-channel) haptic sensor and actuator to enhance the interaction between remote and physical environments. For example, a multi-actuator vibrotactile display utilizing a 2D array of piezoelectric actuators is designed in [2] and spherical tangible objects [3]. Another study in [4] highlighted the multi-actuator haptic gloves for various applications, including virtual reality and teleoperation.

Storing and sending such multi-channel vibrotactile signals requires a large data volume with an increase in the number of channels [5], [6]. Vibrotactile signal compression [5], [7], [8], [9], [10], [11], [12] is designed to represent high-quality signals with a small data volume for single-channel and multi-channel vibrotactile signals. In particular, this study focuses on offline-acquired vibrotactile signals, which are crucial in real-world applications such as virtual training systems, remote robot control in structured environments, or pre-programmed haptic experiences in VR applications. The existing studies [5], [7], [8], [9], [10] aimed to remove perceptual redundant information to realize a compact representation of vibrotactile signals. Here, perceptual redundancy [13] means that components of the signal are not perceived by a user and can be omitted.

Existing studies have proposed signal processing-based solutions to reduce perceptual redundancy. A key concept of the existing solutions is to introduce the psychohaptic model [7], i.e., unequal frequency sensitivity of the haptic feedback considering user perception, for the compression. In the signal processing-based solutions, the state-of-the-art for single-channel and multi-channel vibrotactile signals is Vibrotactile Signal Compression based on Perceptual Wavelet Quantization (VC-PWQ) [8] and MVibCode [5]. In both methods, each channel of the vibrotactile signal is transformed into a frequency-domain representation using discrete Wavelet transform (DWT). A different amount of bits is assigned to each band of DWT coefficients based on the psychohaptic model to remove the perceptual redundancy. Here, MVibCode further introduces the clustering method for the multi-channel vibrotactile signals to realize the differential compression across the channels. However, existing studies on single-channel haptic codecs using neural network (NN) demonstrated that the typical signal processing-based codecs suffered from low coding efficiency [11].

Our study aims to obtain a compact representation of multi-channel vibrotactile signals empowered by NN. For this

purpose, we propose a novel scheme for multi-channel vibrotactile compression inspired by the concept of implicit neural representation (INR) [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25]. INR architectures consist of a small NN architecture to train the coordinate-to-value mappings. It overfits the desired signal to a small NN architecture through the training process, and then takes the feed-forward process for the overfitted NN architecture to reconstruct the desired signal. Although the overfitting process requires a large encoding delay, the proposed scheme exploits the concept of INR for a memory-efficient format by overfitting the desired multi-channel vibrotactile signals to the proposed NN architecture and regards the overfitted weights of the NN architecture as a compact representation of multi-channel vibrotactile signals. We additionally design a novel loss function inspired by the psychohaptic model to yield a compact representation considering the perceptual redundancy.

Experiments using a multi-channel vibrotactile dataset [26] and perception-aware metric [27] show that the proposed scheme can obtain compact representation from the multi-channel vibrotactile signals with less perceptual quality degradation compared to the state-of-the-art vibrotactile compression schemes. In addition, the proposed psychohaptic-inspired loss function brings further compact representation under the same perceptual quality against the typical mean square error (MSE) loss function.

Related Work and Contributions: In recent years, an auto-encoder (AE)-based solution [11] has been proposed for another NN-empowered vibrotactile compression. In contrast to the typical AE-based compression studies for image and video signals, it adopts a loss function considering perceptual threshold with the damage threshold inspired by the psychohaptic model for training encoding and decoding architecture to yield compact representation considering user perception from the vibrotactile signals.

A key difference between the AE-based architecture and the INR architecture is to train the architecture using numerous vibrotactile signals or single vibrotactile signals. It brings an advantage even when the training dataset is limited.

The existing INR architectures [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25] are mainly designed for image and video signals. Our study is the first to propose an INR-empowered vibrotactile compression. For efficient representation of multi-channel vibrotactile, the proposed scheme has three-fold contributions.

- The proposed scheme decodes the desired channel's vibrotactile signals by feeding the channel ID to the INR architecture.
- A loss function considering a psychohaptic model to obtain a memory-efficient format considering user perception.
- We use an open multi-channel vibrotactile dataset for experiments and evaluate the compression rate (CR) and perceptual quality in terms of Spectral Temporal SIMilarity (ST-SIM) [27] across different user's gestures.

II. Proposed Scheme

A. Overview

Fig. 1 shows an end-to-end architecture of the proposed scheme. The proposed scheme assumes a scenario in which N three-axis accelerometers [28], designed to record vibrotactile signal, are attached to the fingers. Each accelerometer is assigned a unique ID i. Each accelerometer records the three-dimensional vibrotactile signal, which is transformed into a one-dimensional vibrotactile signal using DFT321 [29]. Here, the one-dimensional vibrotactile signal corresponding to accelerometer i is defined as $\boldsymbol{v}_i \in \mathbb{R}^T$, where T represents the number of samples in the one-dimensional vibrotactile signal. The vibrotactile signals collected from multiple accelerometers are integrated into a multi-channel vibrotactile signal V, defined as:

$$V = \{\boldsymbol{v}_i \mid i = 1, \ldots, N\}.$$

Based on the multi-channel vibrotactile signal, we construct a dataset $\mathcal{D}$, which is a set of the tuple of each normalized ID $i \in [0, 1]$ along with its corresponding ID's vibrotactile signal $\boldsymbol{v}_i$, as follows:

$$\mathcal{D} = \{(i, \boldsymbol{v}_i) \mid i \in [0, 1]\}.$$

In the encoder-side, an INR architecture is defined as $\Phi(\cdot; \boldsymbol{\theta}) : \mathbb{R}^1 \rightarrow \mathbb{R}^T$ with learnable parameters $\boldsymbol{\theta}$ to overfit the desired multi-channel vibrotactile signal. During the training process, the parameters $\boldsymbol{\theta}$ are optimized to learn the mapping from the ID i to its corresponding vibrotactile signal $\boldsymbol{v}_i$ using the dataset $\mathcal{D}$. For this optimization, a loss function g is adopted to minimize the perceptual loss between the original $\boldsymbol{v}_i$ and the reconstructed vibrotactile signal $\hat{\boldsymbol{v}}_i = \Phi(i; \boldsymbol{\theta})$. In summary, the optimization problem is defined as follows:

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, g(\boldsymbol{v}_i, \Phi(i; \boldsymbol{\theta})) \tag{1}$$

The trained parameters $\boldsymbol{\theta}^*$ are then pruned and quantized to remove redundant information. The compressed parameters are stored or transmitted as a lightweight representation of the multi-channel vibrotactile signal.

In the decoder, the ID i's vibrotactile signal $\hat{\boldsymbol{v}}_i$ is reconstructed by the INR architecture $\Phi(i; \boldsymbol{\theta}^*)$ parameterized by the compressed parameters $\boldsymbol{\theta}^*$. Here, the INR architecture in the decoder is identical to that in the encoder. The accelerometer ID i is fed into the INR in parallel, and the corresponding vibrotactile signal is reconstructed. By integrating the reconstructed vibrotactile signal of all the accelerometers, the decoder finally decodes the multi-channel vibrotactile signals $\hat{V}$.

B. INR Architecture

Fig. 2 illustrates the detailed structure of the proposed INR. The proposed INR consists of an input embedding layer and L upscale blocks.

The input embedding layer, composed of positional embedding and a multilayer perceptron (MLP), transforms the input, i.e., accelerometer ID i, into a high-dimensional feature. The high-dimensional feature enhances the ability to represent

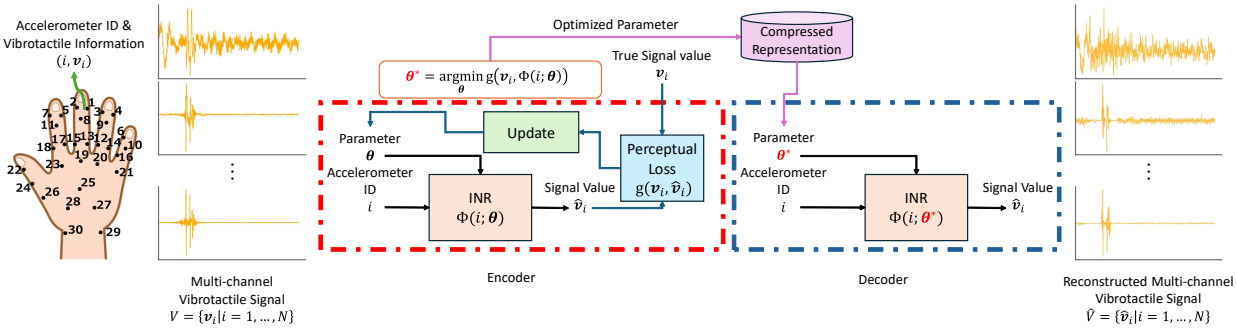


Fig. 1: Overview of the proposed scheme

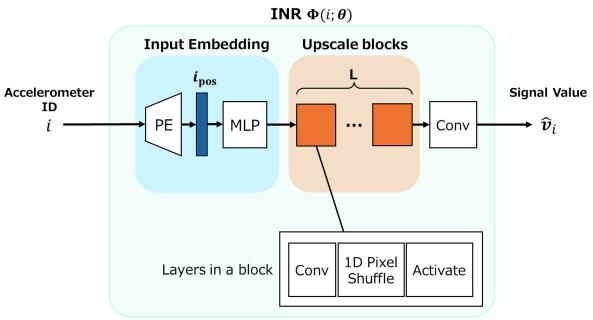


Fig. 2: Architecture of the proposed INR

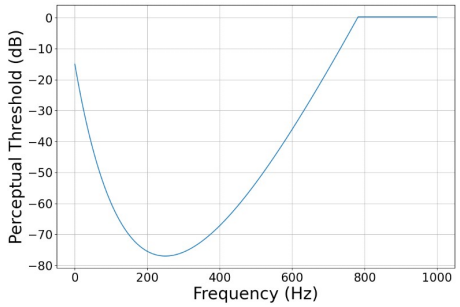


Fig. 3: Perceptual threshold adapted from [8]

high-frequency components of the desired vibrotactile signal under a small neural network.

The upscale block maps the extracted feature to a vector of length T, corresponding to the vibrotactile signal for accelerometer ID i. Each upscale block consists of a one-dimensional convolutional layer, a one-dimensional pixel shuffle layer, and an activation function. Let s_l denote the scaling factor of l-th layer. In the l-th layer, the one-dimensional convolutional layer increases the number of channels in the input feature map w_l to $w_{l-1} \times c \cdot s_l$. The one-dimensional pixel shuffle layer then expands the resolution of the feature map to $w_{l-1} \cdot s_l \times c$. At the L-th upscale block, the resolution of the feature map becomes $w_0 \cdot s_1 \ldots s_L \times c$. In the end, a 1×1 one-dimensional convolutional layer is applied to transform the feature map into a vibrotactile signal with a resolution of $T \times 1$.

C. Loss Function

The mean squared error (MSE) is widely used for training INR architectures as a loss function. The minimization of MSE aims to overfit the INR architecture to the entire signal. However, the loss function does not account for perceptual sensitivity differences within the vibrotactile signal, leading to remaining perceptual redundancy. The proposed scheme introduces a novel loss function based on a psychohaptic model. It considers unequal weight for each frequency band of the vibrotactile signal to achieve a more compact data representation.

The loss function is designed based on the perceptual threshold defined in the psychohaptic model [7]. The perceptual threshold represents the minimum level required for a stimulus of a signal to be perceivable, being defined at each frequency, quantifying human tactile sensitivity across different frequency bands. The psychohaptic model indicates that the perceptual threshold rises sharply beyond 800 Hz, which implies tactile feedback in higher frequency bands is nearly impossible to perceive [8].

However, signals with sufficiently high amplitudes and pressures in high-frequency bands can still cause discomfort or damage to user perception. Thus, the perceptual threshold may fail to reproduce human perceptual characteristics. In the audio domain, the damage threshold is approximately 90 dB higher than the minimum perceptual threshold [30], serving as a benchmark for evaluating the physical effects of strong stimuli on sensory organs. In the tactile domain [8], they consider that the damage threshold is approximately 77 dB higher than the minimum perceptual threshold, where the minimum perceptual threshold is -77 dB.

In our study, we follow the perceptual threshold considering the damage threshold and cut off the frequency components of the vibrotactile signal exceeding 0 dB in the perceptual threshold to effectively reduce the perceptual redundancy within the vibrotactile signal of each accelerometer. The perceptual threshold considering the damage threshold is defined as follows:

$$t(f) = \begin{cases} \left| \frac{62\,\mathrm{dB}}{\left(\log_{10}\left(\frac{6}{11}\right)\right)^2} \left[\log_{10}\left(\frac{f}{550\,\mathrm{Hz}} + \frac{6}{11}\right)\right]^2 \right| - 77\,\mathrm{dB}, & \text{if } 0 \le f \le 780 \\ 0, & \text{if } f > 780 \end{cases} \tag{2}$$

Fig. 3 shows the perceptual threshold considering the damage threshold as a function of frequency bands. It indicates that users more easily perceive vibrotactile responses below 780 Hz. In addition, the haptic feedback in high frequency bands may cause some effects in user perception, and thus such components are considered as "Do not care" $\emptyset$ for reconstruction. Based on this property, we define a filter $h(f)$

to cut off the contribution of high-frequency components as follows:

$$h(f) = \begin{cases} 1, & \text{if } 0 \leq f \leq 780 \\ \emptyset, & \text{if } f > 780 \end{cases} \tag{3}$$

Our loss function is designed to compare the frequency components of the original and the reconstructed vibrotactile signals using the binary filter. Let $S_i(f)$ denote f-th frequency response obtained by applying the discrete cosine transform (DCT) to i-th original vibrotactile signal $\boldsymbol{v}_i$, and $\hat{S}_i(f)$ denote the frequency response obtained from the i-th reconstructed vibrotactile signal $\Phi(i;\boldsymbol{\theta})$. The loss function is defined as:

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{i=1}^{N} g(\boldsymbol{v}_i, \hat{\boldsymbol{v}}_i) \tag{4}$$

$$g(\boldsymbol{v}_i, \hat{\boldsymbol{v}}_i) = \frac{\sum_f \left(h(f) \cdot S_i(f) - h(f) \cdot \hat{S}_i(f)\right)^2}{\sum_f \left(h(f) \cdot S_i(f)\right)^2} \tag{5}$$

The loss function emphasizes the response gap in frequency bands with high sensitivity, effectively preserving perceptually important components while optimizing data representation.

D. Model Compression

The trained parameters $\boldsymbol{\theta}^*$ efficiently represent multi-channel vibrotactile signals. To further enhance compactness, we introduce a sequence of model compression operations.

1) Model Pruning: In the beginning, global unstructured pruning is applied to the trained parameters. Specifically, each parameter $\theta_i \in \boldsymbol{\theta}^*$ is retained or pruned based on the magnitude of the parameter $|\theta_i|$ and a threshold θ_q based on the q-percentile magnitude of the parameters as follows:

$$\theta = \begin{cases} \theta_i, & \text{if } \theta_i \geq \theta_q \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

After the model pruning, the INR architecture is retrained and fine-tuned using the same dataset D to retain adequate representational capability using the pruned parameters.

2) Model Quantization and Encoding: The pruned and fine-tuned parameters are uniformly quantized to a bit depth N_b. This quantization is a layer-wise operation. For a parameter set $\mu \in \boldsymbol{\theta}$ corresponding to each layer of the INR architecture, the quantized parameter set μ_q is obtained as follows:

$$\mu_q = \operatorname{round}\left(\frac{\mu - \mu_{\min}}{2^{N_b}}\right) s + \mu_{\min},\ s = \frac{\mu_{\max} - \mu_{\min}}{2^{N_b}}, \tag{7}$$

where $\operatorname{round}(\cdot)$ denotes rounding to the nearest integer, and $\mu_{\max}$ and $\mu_{\min}$ represent the maximum and minimum values of the parameters $\boldsymbol{\mu}$, respectively. The quantized tensor μ_q is then converted into a binary sequence using Huffman coding. Notably, for smaller bit depths, the quantized parameters μ_q are more likely to take values near zero, making Huffman coding particularly effective for compression.

III. Evaluation

A. Settings

Datasets: An open vibrotactile dataset provided by [26] was used for evaluation. The dataset consists of touch-evoked skin vibrations recorded by 30 accelerometers from four subjects. The accelerometers were placed on the dorsal surface of the subject's right hand to prevent interference with touch interactions. Due to the characteristics of the vibrations in the hand, the recorded signals are very similar to those on the volar surface. Vibrotactile signals were measured while subjects performed 13 different manual gestures. The gestures were chosen to be similar to those used when interacting with the environment in everyday life. Specifically, in gestures 1 to 5, a subject taps a steel plate using the thumb, index finger, middle finger, ring finger, and little finger, respectively. Gesture 6 represents the tapping action on the same steel plate using both the index and middle fingers, while gesture 7 is a tapping action performed with the index, middle, ring, and little fingers. Gesture 8 is a tapping action using all fingers. In gesture 9, a subject slides the index finger across the surface of a steel plate to perceive its texture. In gestures 10 and 11, a subject grasps cylindrical objects using the thumb and index finger. Gesture 12 is the action of grasping a plastic sphere using all fingers. Finally, Gesture 13 represents a tapping action on the surface of a steel plate using a stylus held between the thumb and index finger.

This paper selected vibrotactile signals from one of the four subjects for evaluation. In addition, this paper initially used the measured vibrotactile signals when a subject taps a steel plate with the ring finger, i.e., gesture 4, to evaluate the baseline performance and the other 12 gestures to discuss the gesture-invariant performance of the baseline schemes.

Metric: We use CR as the metric for the data size representing multi-channel vibrotactile signals. CR is defined as follows:

$$\text{CR} = \frac{\text{Total number of bits for original signals}}{\text{Total number of bits for compressed representation}} \tag{8}$$

In the original vibrotactile signal, 32 bits are assigned per sample, and 1200 samples are measured using each accelerometer.

Two metrics of the vibrotactile signal quality were considered for comparison: (1) ST-SIM [27] and (2) Signal-to-Noise Ratio (SNR) [31]. SNR is defined as follows:

$$\text{SNR} = 10 \log_{10}\left(\frac{\|V\|_F^2}{\|\hat{V} - V\|_F^2}\right), \tag{9}$$

where $\|V\|_F^2$ is the squared Frobenius norm of the matrix V. ST-SIM is a perceptual metric for vibrotactile signals. ST-SIM uses spectral and temporal properties of the original and reconstructed vibrotactile signals to compute a perceptual similarity score between 0 and 1.

Baseline: The first and second baselines are VC-PWQ [8] and MVibCode [5], respectively. VC-PWQ is the state-of-the-art for vibrotactile compression and a modified version of haptic codec [7]. MVibCode extends VC-PWQ to multi-channel vibrotactile signals. Specifically, it proposes a clustering module

over multiple accelerometers to find the cluster head and implements differential coding between the cluster head and cluster members in each cluster. We set a clustering threshold g_{thr} of 0.1 to show the best performance of MVibCode with the clustering module. Note that the performance of MVibCode and VC-PWQ is the same at $g_{\mathrm{thr}} = 0$, since it does not use clustering between the multiple channels.

The other baselines are variants of the proposed schemes under the different loss functions. The first variant is the proposed scheme with a loss function of MSE between the original and reconstructed vibrotactile signals. The improvement against this baseline represents the quality improvement considering the perceptual threshold. The second variant integrates the MSE and proposed loss functions to discuss the impact of the integration on the reconstruction quality.

Hyperparameter Details: We train an individual INR model for each gesture using the Adam optimizer at a learning rate of 1e-3. We adopt a cosine annealing learning rate schedule, with a batch size of 1, 500 training epochs, and 100 warmup epochs. The INR architecture consists of 4 upscale blocks with upscale factors of 10, 8, 5, and 3, respectively. By varying the hidden dimensions of the MLP and the channel dimensions of the upscale blocks, we can construct INR models of different sizes. For input embedding parameters, we set $b = 1.25$ and $l = 30$.

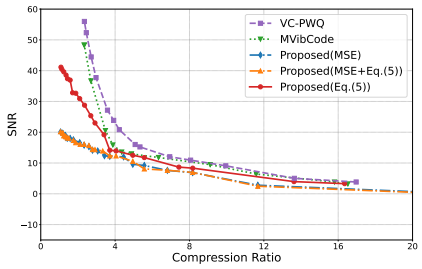
Implementation Detail: All the evaluations are performed with CPUs of Intel Core i9-10850K and i9-13900KF and with GPUs of NVIDIA GeForce RTX 3080. Our proposed method is implemented, trained, and evaluated using PyTorch 2.2.0 with Python 3.10. Here, the average decoding latency of the proposed scheme is 2.219 ms (approximately, 676 kHz). Two baselines were implemented using MATLAB software.

B. Baseline Performance

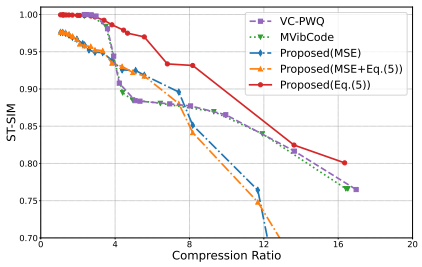
Fig. 4 (a) shows the quality of the reconstructed vibrotactile signals in terms of SNR under the different CRs. We can see the following key observations:

- VC-PWQ achieves the best SNR performance, especially in low CR regimes.
- The proposed scheme achieves the best performance against the variants of other proposed schemes under the different CRs.
- Other proposed schemes are less quality under the same CRs since training the entire signal using a small NN architecture is difficult.
- MVibCode with the clustering module suffers from lower reconstruction quality than the SNR index of VC-PWQ, irrespective of CRs. MVibCode is well-performed in a closed multi-channel vibrotactile signal dataset [5]. However, the clustering-based differential coding causes low quality at high CR regimes using an open dataset [26].

Fig. 4 (b) shows the ST-SIM index under the different CRs. We can see that the proposed scheme achieves the best perceptual quality when the CR is up to 16. The other variants of the proposed scheme do not achieve a similar perceptual quality. VC-PWQ follows almost the same quality in a low



(a) SNR



(b) ST-SIM

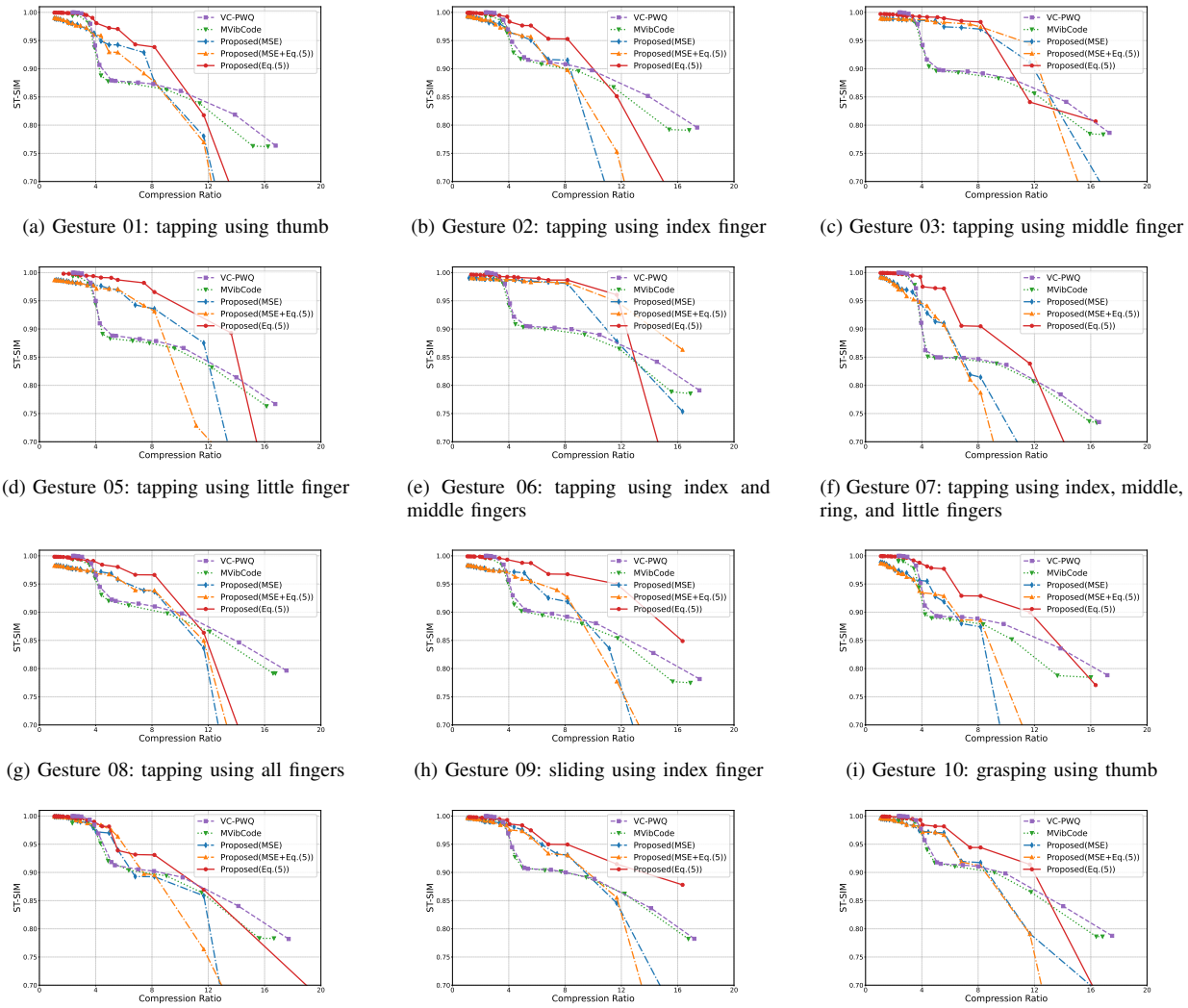
Fig. 4: Reconstruction quality as a function of compression ratio under the different quality metrics in gesture 04. Here, the gesture is a tapping action on a steel plate using the ring finger. (a) SNR, (b) ST-SIM

CR regime. However, the perceptual quality suddenly degrades at a certain CR because the bit depth for the mid-frequency components, i.e., those sensitive to human perception, must be reduced as the CR increases. The above results highlight that 1) the proposed representation is effective in storing/sending high-quality, e.g., ST-SIM index above 0.9, and low-rate multi-channel vibrotactile signals compared to the state-of-the-art vibrotactile coding solutions, and 2) the proposed loss function considering the perceptual threshold contributes to improving the perceptual quality compared to the typical MSE loss function. Here, the ST-SIM score is linearly correlated with the subjective ratings, i.e., the mean opinion score (MOS) [27], and thus even a 0.1 improvement in ST-SIM results in an improvement in the perceptual quality of the vibrotactile experience.

C. Impact of User Gestures

Figs. 5 (a) through (l) show the results of the ST-SIM index as a function of CRs for 12 different gestures. The detailed setting of each gesture is mentioned in Sec. III-A. We can see the following findings from the results:

- When we see the gap between the proposed scheme and VC-PWQ, the proposed scheme maintains a high ST-SIM index, i.e., above 0.9, in a high CR regime compared to VC-PWQ, regardless of the gestures.
- In some gestures, the ST-SIM index of the proposed scheme suddenly degrades at a high CR, and it is lower than the ST-SIM index of VC-PWQ and MVibCode.



(a) Gesture 01: tapping using thumb

(b) Gesture 02: tapping using index finger

(c) Gesture 03: tapping using middle finger

(d) Gesture 05: tapping using little finger

(e) Gesture 06: tapping using index and middle fingers

(f) Gesture 07: tapping using index, middle, ring, and little fingers

(g) Gesture 08: tapping using all fingers

(h) Gesture 09: sliding using index finger

(i) Gesture 10: grasping using thumb

(j) Gesture 11: grasping using index finger

(k) Gesture 12; grasping using all fingers

(l) Gesture 13: tapping using a stylus held between thumb and index finger

Fig. 5: ST-SIM index as a function of compression ratios across different gestures. We note that the performance in gesture 04 was shown in Fig.4 (b).

- In gestures 3 and 6, an integration of MSE and proposed loss functions maintains a high ST-SIM index at a high CR.

Among the gestures, gesture 09 (sliding using the index finger) revealed the most significant improvement in ST-SIM over the baselines, demonstrating the effectiveness of the proposed scheme for continuous tactile movements. In contrast, gesture 08 (tapping using all fingers) showed the lowest improvement, suggesting that simultaneous multi-finger activations may introduce complex interference patterns that challenge the compact representation.

IV. Conclusion and Future Work

This paper proposed a novel coding scheme for multi-channel vibrotactile signals. The proposed scheme integrated an INR architecture for training accelerometer ID-to-signal mapping, a loss function based on a psychohaptic model, and model compression for overfitted weights to yield a compact representation of vibrotactile signals with a slight degradation of perceptual quality. Experiments demonstrated that the proposed scheme realizes better-quality vibrotactile signals under the same CR, especially up to a CR of 10, compared with the VC-PWQ and MVibCode schemes across different gestures.

There are several extensions for the INR architecture and the loss function for future work. For example, Fourier-based [32], [33], [34] and Wavelet-based [35] embedding can represent high-frequency details in a small model size. In addition, an optimal integration of MSE and proposed loss functions with the entropy function [36] can achieve the best rate-distortion performance for each subject and corresponding gesture.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number JP22H03582.

REFERENCES

[1] M. Emami, A. Bayat, R. Tafazolli, and A. Quddus, "A survey on haptics: Communication, sensing and feedback," *IEEE Communications Surveys & Tutorials*, vol. PP, no. 99, pp. 1–45, 2024.

[2] L. Pantera and C. Hudin, "Multitouch vibrotactile feedback on a tactile screen by the inverse filter technique: Vibration amplitude and spatial resolution," *IEEE Transactions on Haptics*, vol. 13, no. 3, pp. 493–503, 2020.

[3] P.-A. Cabaret, T. Howard, G. Gicquel, C. Pacchierotti, M. Babel, and M. Marchal, "Does multi-actuator vibrotactile feedback within tangible objects enrich vr manipulation?" *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, no. 8, p. 4767–4779, 2023.

[4] S. M, K. Venusamy, S. S, S. S, and N. K. O, "A comprehensive review of haptic gloves: Advances, challenges, and future directions," in *Second International Conference on Electronics and Renewable Systems*, 2023, pp. 227–233.

[5] L. Nockenberg, A. Noll, S. Panëels, A. B. Dhiab, C. Hudin, and E. Steinbach, "Mvibcode: Multi-channel vibrotactile codec using hierarchical perceptual clustering," *IEEE Transactions on Haptics*, vol. 16, no. 4, pp. 646–651, 2023.

[6] T. Ogura, S. Kitamura, T. Fujihashi, S. Saruwatari, and T. Watanabe, "Joint source-channel coding for multi-channel vibrotactile sensors," *IEEE Access*, vol. 13, pp. 7736–7745, 2025.

[7] A. Noll, B. Gülecyüz, A. Hofmann, and E. Steinbach, "A rate-scalable perceptual wavelet-based vibrotactile codec," in *IEEE Haptics Symposium*, 2020, pp. 854–859.

[8] A. Noll, L. Nockenberg, B. Gülecyüz, and E. Steinbach, "VC-PWQ: vibrotactile signal compression based on perceptual wavelet quantization," in *IEEE World Haptics Conference (WHC)*, 2021, pp. 427–432.

[9] R. Chaudhari, C. Schuwerk, M. Danaei, and E. Steinbach, "Perceptual and bitrate-scalable coding of haptic surface texture signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 3, pp. 462–473, 2015.

[10] R. Hassen, B. Gülecyüz, and E. Steinbach, "PVC-SLP: perceptual vibrotactile-signal compression based-on sparse linear prediction," *IEEE Transactions on Multimedia*, vol. 23, pp. 4455–4468, 2021.

[11] L. Nockenberg and E. Steinbach, "Vibrotactile signal compression using perceptually trained autoencoders," in *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer, 2024, pp. 264–277.

[12] Y. Xu, D. Chen, Y. Fang, Y. Lu, and T. Zhao, "Efficient vibrotactile codec based on nbeats network," *IEEE Signal Processing Letters*, vol. 31, pp. 2845–2849, 2024.

[13] Y. Zhang, T. Wang, Y. Xiao, T. Zhang, Y. Zhang, and R. Tao, "SJ-PVC: An efficient perceptual video compression scheme based on adaptive qp and rate-distortion optimization," *IEEE Transactions on Consumer Electronics*, vol. PP, no. 99, pp. 1–14, 2025.

[14] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," *Advances in neural information processing systems*, vol. 33, pp. 7462–7473, 2020.

[15] E. Dupont, A. Golinski, M. Alizadeh, Y. W. Teh, and A. Doucet, "COIN: COmpression with implicit neural representations," in *Neural Compression: From Information Theory to Applications – Workshop*, 2021.

[16] E. Dupont, H. Loya, M. Alizadeh, A. Goliński, Y. W. Teh, and A. Doucet, "COIN++: neural compression across modalities," *Transactions on Machine Learning Research*, vol. 2022, no. 11, pp. 1–26, 2022.

[17] H. Chen, B. He, H. Wang, Y. Ren, S.-N. Lim, and A. Shrivastava, "NeRV: Neural representations for videos," in *Advances in Neural Information Processing Systems*, 2021.

[18] H. M. Kwan, G. Gao, F. Zhang, A. Gower, and D. Bull, "HiNeRV: Video compression with hierarchical encoding-based neural representation," in *Advances in Neural Information Processing Systems*, 2023.

[19] J. C. Lee, D. Rho, J. H. Ko, and E. Park, "FFNeRV: Flow-guided frame-wise neural representations for videos," in *Proceedings of the ACM International Conference on Multimedia*, 2023, p. 7859–7870.

[20] B. He, X. Yang, H. Wang, Z. Wu, H. Chen, S. Huang, Y. Ren, S.-N. Lim, and A. Shrivastava, "Towards scalable neural representation for diverse videos," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023.

[21] Y. Xu, X. Feng, F. Qin, R. Ge, Y. Peng, and C. Wang, "VQ-NeRV: A vector quantized neural representation for videos," *arXiv e-prints*, Mar. 2024.

[22] S. R. Maiya, S. Girish, M. Ehrlich, H. Wang, K. Lee, P. Poirson, P. Wu, C. Wang, and A. Shrivastava, "Nirvana: Neural implicit representations of videos with adaptive networks and autoregressive patch-wise modeling," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, jun 2023, pp. 14 378–14 387.

[23] R. Xue, J. Li, T. Chen, D. Ding, X. Cao, and Z. Ma, "NeRI: Implicit neural representation of lidar point cloud using range image sequence," 2024, pp. 8020–8024.

[24] Y. Bai, C. Dong, C. Wang, and C. Yuan, "PS-NeRV: Patch-wise stylized neural representations for videos," in *IEEE International Conference on Image Processing*, 2023, pp. 41–45.

[25] H. Yan, Z. Ke, X. Zhou, T. Qiu, X. Shi, and D. Jiang, "DS-NeRV: Implicit neural video representation with decomposed static and dynamic codes," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 23 019–23 029.

[26] Y. Shao, V. Hayward, and Y. Visell, "Compression of dynamic tactile information in the human hand," *Science Advances*, vol. 6, no. 16, pp. 1–8, 2020.

[27] R. Hassen and E. Steinbach, "Subjective evaluation of the spectral temporal similarity (ST-SIM) measure for vibrotactile quality assessment," *IEEE Transactions on Haptics*, vol. 13, no. 1, pp. 25–31, 2020.

[28] J. Kirsch, A. Noll, M. Strese, Q. Liu, and E. Steinbach, "A low-cost acquisition, display, and evaluation setup for tactile codec development," in *2018 IEEE International Symposium on Haptic, Audio and Visual Environments and Games (HAVE)*. IEEE, 2018, pp. 1–6.

[29] N. Landin, J. M. Romano, W. McMahan, and K. J. Kuchenbecker, "Dimensional reduction of high-frequency accelerations for haptic rendering," in *International conference on human haptic sensing and touch enabled computer applications*. Springer, 2010, pp. 79–86.

[30] M. E. Lutman, "What is the risk of noise-induced hearing loss at 80, 85, 90 db(a) and above?" *Occupational Medicine*, vol. 50, no. 4, pp. 274–275, 05 2000.

[31] X. Liu, M. Dohler, and Y. Deng, "Vibrotactile quality assessment: Hybrid metric design based on SNR and SSIM," *IEEE Transactions on Multimedia*, vol. 22, no. 4, pp. 921–933, 2020.

[32] N. Benbarka, T. Höfer, A. Zell *et al.*, "Seeing implicit neural representations as Fourier series," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 2041–2050.

[33] D. Xu, P. Wang, Y. Jiang, Z. Fan, and Z. Wang, "Signal processing for implicit neural representations," *Advances in Neural Information Processing Systems*, vol. 35, pp. 13 404–13 418, 2022.

[34] M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. Barron, and R. Ng, "Fourier features let networks learn high frequency functions in low dimensional domains," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7537–7547, 2020.

[35] V. Saragadam, D. LeJeune, J. Tan, G. Balakrishnan, A. Veeraraghavan, and R. G. Baraniuk, "Wire: Wavelet implicit neural representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 18 507–18 516.

[36] C. Gomes, R. Azevedo, and C. Schroers, "Video compression with entropy-constrained neural representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 18 497–18 506.