OBJECTIVE QUALITY PREDICTION FOR HAPTIC TEXTURE SIGNAL COMPRESSION

Rahul Chaudhari[⋆] *Yongjae Yoo*[†] *Clemens Schuwerk*[⋆] *Seungmoon Choi*[†] *Eckehard Steinbach*[⋆]

[⋆] Chair of Media Technology, Technische Universität München, Germany [*]
[†] Department of Computer Science & Engineering, POSTECH, Republic of Korea [†]

ABSTRACT

Perceptual quality for media compression algorithms is traditionally evaluated through user studies. Such studies are time-consuming, laborious and expensive, slowing down the development of new signal processing algorithms. To address this problem, a number of algorithmic quality prediction methodologies have been developed in the audio and video fields, something that is currently lacking in haptics research.

In this paper, we present a novel method for predicting the perceptual quality degradation of compressed haptic texture signals. For this purpose, abstract perceptual features like Roughness, Brightness, etc. that capture the subjective experience of textures are exploited, in addition to low-level psychophysical models from the literature. As compared to the state-of-the-art, the presented prediction methodology shows an approximately 30% improvement in explaining the variance in the perceptual data.

***Index Terms*—** data compression, haptic signals, subjective/objective quality

1. INTRODUCTION

Certain haptic properties of a textured surface can be captured by scanning it with a tool and recording the generated wideband acceleration signals (Slave-side of Fig. 1). Displaying these signals to the human through a *vibrotactile* actuator (Master-side of Fig. 1) improves the perceived haptic realism and task performance in teleoperation scenarios [1, 2].

Efficient communication of haptic data over a network in a teleoperation system necessitates data compression [3]. Lossy data compression may lead to perceivable distortion in the haptic signals, which needs to be evaluated via user studies. User studies are usually difficult and time-consuming. In this paper, we seek to eliminate the need for such user studies by developing an algorithm for Objective Quality Prediction (OQP) for compressed haptic signals.

In this context, we study the distortion introduced by the Code-Excited Linear Predictor-based haptic texture codec presented in [3]. In this codec, a linear predictive (LP) filter captures the spectral envelope of the haptic texture signal, while a pair of codebooks capture the salient time-domain features. The LP and codebook parameters are vector quantized before transmission to the receiver side. Starting from an input bitrate of 32 kbps, this codec has been used to generate five compressed bitrates (3.15, 2.75, 2.45, 2.15, and 1.75 kbps), corresponding to five perceptual quality degradations for the experiments performed in the present paper.

The remainder of this paper is organized as follows. We first review the state-of-the-art for haptic texture OQP. Next, user studies to generate ground-truth perceptual data for compressed texture signals are described in Section 2. Section 3 presents a novel OQP method to predict the perceptual data based on a linear combination of haptic perceptual features. We conclude the paper in Section 4 with ideas for future work.

Fig. 1: A haptic teleoperation system. The user commands the Slave robot's movements through the Master robot. The acceleration signals captured at the interface of the Slave robot and the object surface are transmitted back to the Master side and displayed to the user's hand through a vibrotactile actuator.

Related work - OQP for compressed haptic data

In [4], Okamoto and Yamada present an objective metric for quantifying perceptual quality degradation caused by haptic texture signal compression. Their metric is based on a spectral model for vibrotactile perception mediated by Pacinian corpuscles in the human hand [5].

In [5], Bensmaïa et al. hypothesize the presence of frequency-domain Gaussian-shaped "minichannels" mediating vibrotactile perception, and calculate the "activation" generated in each of these minichannels by a given stimulus. The perceptual dissimilarity $D_{S_1 S_2}$ between two stimuli S_1 and S_2 is estimated by taking the difference between the ac-

*This work has been supported by the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Grant agreement no. 258941.

†This work was supported by the NRF of Korea grant No. 2013R1A2A2A01016907.

tivations Z_{S_1} and Z_{S_2} they generate in the minichannels, and summing up the absolute differences across all minichannels: $D_{S_1S_2} \propto \sum_{f_c} |Z_{S_1}(f_c) - Z_{S_2}(f_c)|$, where f_c's are the center frequencies of 100 minichannels, placed in logarithmic increments in the Pacinian frequency range (40-1000 Hz), and minichannel activation

$$Z_S(f_c) = \sum_f \left(\frac{A_f^2}{T_f^2}\right)^{a_f} \cdot e^{-\frac{(f-f_c)^2}{2(\alpha f_c)^2}},$$

where frequency f lies in (40-1000 Hz), A_f is the signal amplitude and T_f the Pacinian detection threshold, both as functions of frequency; α controls the spread of the gaussian filter.

The work in [4] supplements the above power-spectral model with an analogous amplitude-spectral model (replacing (A_f^2/T_f^2) with (A_f/T_f)). This change accounts for Meissner corpuscle-mediated lower frequency vibrations (tens of Hz), which were ignored due to the inverse of the square of the (U-shaped) T_f function. With this extension, the algorithm in [4] is able to predict perceptual dissimilarity for compressed texture signals with a drastically improved Goodness-of-Fit (R^2 = 0.64, up from 0.2 for our texture data).

The above two models capture low-level perceptual features of texture signals by focusing only on one particular aspect of vibrotactile perception - the Pacinian/Meissner detection thresholds. In this paper, we additionally use high-level perceptual features - Roughness [6], Brightness [8], Regularity, time-envelope pattern [9], etc. - to predict perceptual quality data. We then combine and weight all these features to best predict perceptual degradation (dissimilarity) data for texture compression.

2. PERCEPTUAL QUALITY EVALUATION

In this section, we describe psychophysical experiments conducted to generate ground-truth perceptual dissimilarity (degradation) data for the codec input-output signals.

2.1. Subjects

Eight right-handed subjects (2 female, 6 male) with ages between 22 to 27 years volunteered for the experiments. None of them reported having any sensory ailments that would affect their perceptual performance.

2.2. Stimuli

Texture signals were recorded with an accelerometer-mounted stylus resting on a textured surface while the surface rotated at a constant speed. The pressing force of the probe on the surface was also controlled to be constant. Five textures varying widely in material and surface patterns were used - a steel texture, a rubber pad, hard wood, leather, and marble. Each texture signal was then processed with the bitrate-scalable codec from [3]. This resulted in five distorted output signals per input reference signal, corresponding to five (decreasing) bitrates (see Section 1). The input-output signal-pairs were divided into 3 segments of 0.8 s each. Fig. 2 shows input-output signals at three distortion levels for the steel texture. Each pair of input-output segments constitute a separate stimulus pair.

Trials with ground-truth answer “different” presented the reference and the distorted stimuli within each pair to the subject in succession, with a gap of 0.6 s. Additionally, trials with ground-truth answer “same” were presented where the reference stimulus was repeated twice. The subject then indicated if he perceived the two stimuli to be the “same” or “different”.

The stimulus duration was chosen to be 0.8 s to avoid adaptation to the stimulus. The pause duration was chosen to be 0.6 s, to avoid haptic enhancement and summation effects [10] in the display of consecutive stimuli. Each trial thus lasted about 2.2 s, which falls within the constraint of several seconds imposed by the haptic working memory [11]. This was followed by a minimum 1.3 s pause, before the subject could continue with the next trial.

2.3. Experimental setup

A custom-made stylus-like handle similar to that of the PHANToM Omni (Fig. 1) haptic device was mounted on the K2004E01 minishaker (see Fig. 3). The subjects were instructed to hold the stylus like a pen in their dominant hand with a natural 3-finger grip. Their elbow and forearm rested on a wooden plank, which supported a natural wrist position. Subjects wore acoustic noise-canceling headphones (Bose QuietComfort 15) that played pink noise to mask out auditory cues from the experimental apparatus. A cardboard barrier visually shielded the minishaker from the subject. This ensured that the subject only had a haptic contact with the minishaker. Subjects interacted with the experiment GUI through a keyboard.

2.4. Method

As mentioned in Section 2.2, subjects were asked to give binary “same/different” judgments for each pair of texture stimuli presented. Each stimulus pair was repeated 6 times to have higher reliability of the results. A subject could test each stimulus pair as many times as required before making a decision, and moving on to the next pair. She/he tested 180 pairs per texture (3 segments $\times$ 5 bitrate settings per segment $\times$ 6 repetitions per bitrate setting $\times$ 2 (same or different)). About 20 min. were needed at an average for a given texture (about 4 s per trial, and periodic rests).

2.5. Results

For a given stimulus-pair, all subject responses are aggregated, and the number of “hits” (h) and “false alarms” (f) are

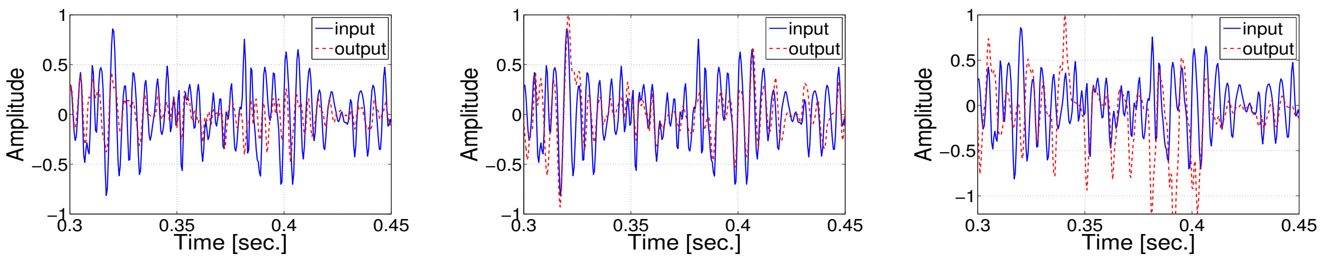

Fig. 2: Three distortion levels (left to right: 3.15, 2.45, and 1.75 kbps) for a segment of the "steel" texture signal. The distorted output segment (- - -) is shown overlapped on the input reference segment (—). Signals are zoomed in for clarity.

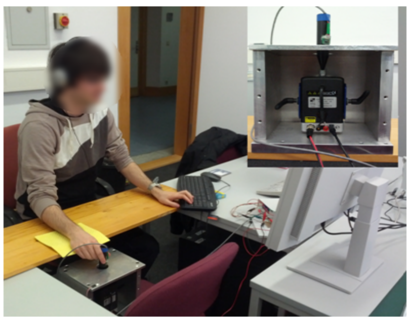

Fig. 3: Experimental setup (minishaker device inset).

counted:

$$h = \sum(\text{decision "different"} \mid \text{ground truth "different"})$$
$$f = \sum(\text{decision "different"} \mid \text{ground truth "same"})$$

The "hit rate" (H) and the "false-alarm rate" (F) are then calculated as $H = h/n$ and $F = f/n$, where $n = 144$ is the number of ground-truth "same" or "different" pairs (3 segments $\times$ 6 repetitions/segment/subject $\times$ 8 subjects) for a given texture signal and bitrate setting.

For our analyses, we employ concepts from Signal Detection Theory [12]. Specifically, we compute a parameter called d' that represents perceptual distance between stimuli belonging to two different classes. d' is occasionally refered to as "dissimilarity" in this paper, similar to [5]. Low values of d' indicate that the stimuli are not easily discriminable, whereas high values indicate higher discriminability. d' is calculated as $d' = z(H) - z(F)$, where $z()$ denotes the inverse Gaussian cumulative distribution function (assuming a Gaussian noise model for the human sensory process).

Figure 4 shows the d' results for all five textures. It can be seen that d' does not always monotonically increase as the bitrate reduces. This can be explained by the nonlinearity of the vector quantization operating on the LP filter parameters. Moreover, any change in the codebook parameters propagates into the future codec output due to the closed-loop search procedure for the excitation parameters.

3. OBJECTIVE QUALITY PREDICTION

To algorithmically predict the perceptual dissimilarities from the user studies, we extract perceptually meaningful features from the texture signals and combine them within a linear model in this preliminary work. More complex nonlinear models and powerful machine learning techniques may be explored in the future.

A number of tactile features, like the ones identified in [8], could potentially be used - 'slow-fast', 'blunt-sharp', 'bumpy-smooth', 'hard-soft', 'dark-bright', 'thick-thin', 'heavy-light', etc. However, algorithmic quality prediction requires features to have well-defined mathematical models, which are not available in most cases. This limits the spectrum of features that can currently be exploited.

3.1. Feature definitions

A list of the following seven features was compiled, combining ones from the literature and the subjective impressions received from subjects - Roughness [6], Brightness [8], Regularity, stimulus vertical asymmetry, and time-envelope pattern [9]. We also included the Bensmaïa model (Section 1), and the Okamoto amplitudes-spectral model [4] as additional features. The eight feature definitions:

1. Roughness We calculate haptic Roughness as the logarithm of the average signal power, as defined in [6].

2. Spectral centroid "Brightness" The spectral centroid is commonly defined as the "center of mass" of the magnitude spectrum of a signal (sum of frequency-weighted spectral magnitude samples). For haptic textures, higher spectral centroids convey "lively" or "bright" textures.

3. Regularity The regularity of textured surfaces gets reflected in the periodicity of the acceleration signals, which can be captured well with the autocorrelation function.

4. Vertical asymmetry Natural texture scans usually generate acceleration signals that are symmetric about the time-axis. A deviation from such a symmetry causes the hand to be forcefully "pushed" or "pulled" by the haptic device, rather than vibrate - a perceptually very strong and disturbing artifact. To capture such asymmetry we use the standard statistical measure of *skewness* of the signal amplitude samples.

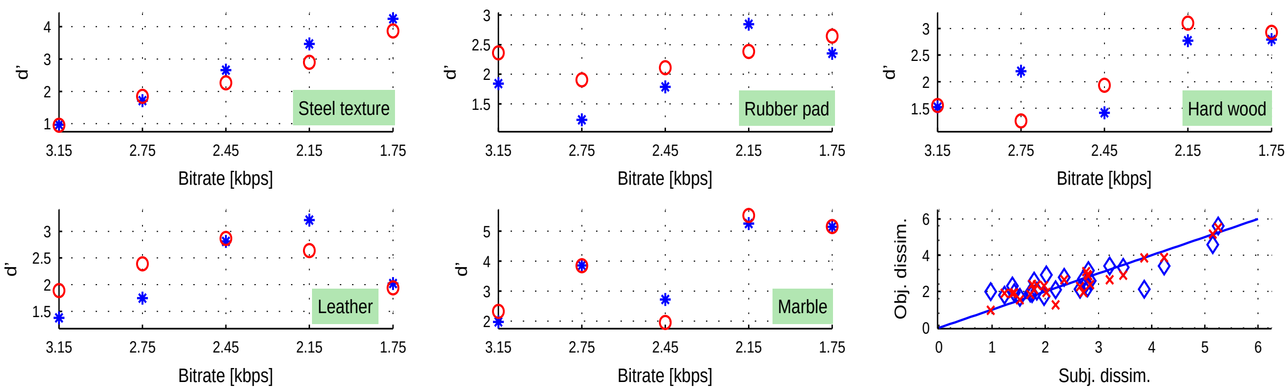

Fig. 4: Objective predictions '○' in comparison with the corresponding subjective results '⋆' for all textures. The right-bottom panel shows a comparison of the prediction performance of all features combined together '×' vs. the Okamoto model [4] alone '◇'.

5. Bensmaïa spectral model and 6. Okamoto model These features, (see Section 1) are based on *perceived* haptic intensity, an important consideration in the development of technical haptic systems [13].

7. Time-envelope Most of the subjects reported memorizing the undulations of a signal as an audio pattern in their minds, and using them for discriminating between two stimuli. This characteristic can be encoded well in the form of time-domain signal envelopes [9]. We use cross-correlation between the time-envelopes of the codec input and output signals as a measure of perceptual distance between them.

3.2. Regression modeling and validation

A linear-regression based feature-selection process filtered the above list down to feature numbers 1, 2, 5, and 7.

A 25×8 feature-distance matrix $\boldsymbol{X}$ was formed for the 25 stimulus pairs (5 textures × 5 distortions). Each matrix contained the Euclidean distance between the input and output stimuli for that feature. The corresponding dissimilarity results from the user studies were stacked in a 25×1 vector $\boldsymbol{y}$.

A feature-selection algorithm ("sequentialfs" Matlab function) was then used to select a minimal subset of features that best predict the data in $\boldsymbol{y}$. The algorithm started by including all eight features at the beginning. Then it dropped features one-by-one until there was noticeable degradation in the prediction performance. Sum of squared linear regression errors was used as the criterion function to select features. Another parameter specified a threshold on the error, thus determining when to stop.

With the reduced feature subset, the data was partitioned into 3 textures for training a linear regression model, and the remaining 2 textures for validation. The textures were exhaustively rotated through the training and validation sets. Regression coefficients for the rotation that led to the least regression error for all textures were selected as final feature weights.

3.3. Results

The objective quality predictions are shown in comparison with the perceptual ground-truth for all textures in Fig. 4. Using only a combination of the Bensmaïa and the Okamoto spectral models, an R^2 (Goodness-of-Fit) of 0.69 is obtained (R^2 adjusted for the number of predictor variables is 0.64). When the features described above were included on top, an improved R^2 of 0.90 (adjusted R^2 0.84) was obtained. Figure 4 right bottom panel shows the superiority of the prediction performance of all the features combined together against that of the Okamoto model alone.

The weights selected for the four features were [roughness: 5.2, brightness: 3.8, spectral: 6.3, time-envelope: 3.3]. It is evident that now, in combination with other features, the Bensmaïa model plays a vital role in explaining the variance in the perceptual data, followed by Roughness, and then the other features.

4. CONCLUSION

This paper presents a novel method for algorithmically predicting perceptual quality degradations for compressed haptic texture signals. We extract perceptually-motivated haptic features that are vital clues for texture signal discrimination, and combine them within a linear model for quality prediction. The results show that abstract features like Roughness, Brightness, and time-envelope pattern play an important role in haptic signal perception, in addition to low-level psychophysical models. A combination of the selected features leads to about 30% improvement over the state-of-the-art in the performance of the quality prediction algorithm.

For future work, one major challenge is to construct mathematical models for haptic perceptual features that have not been accounted for in the present work. Furthermore, the performance of the proposed approach should be investigated for a wider variety of textures.

5. REFERENCES

[1] W. McMahan, J. Romano, A. M. Abdul Rahuman, and K. J. Kuchenbecker, "High frequency acceleration feedback significantly increases the realism of haptically rendered textured surfaces," *Proc. IEEE Haptics Symp.*, pp. 141–148, Mar. 2010.

[2] D. A. Kontarinis and R. D. Howe, "Tactile display of vibratory information in teleoperation and virtual environments," *Presence*, vol. 4, pp. 387–402, 1996.

[3] R. Chaudhari, B. Cizmeci, K. J. Kuchenbecker, S. Choi, and E. Steinbach, "Low bitrate source-filter model based compression of vibrotactile texture signals in haptic teleoperation," *ACM Multimedia*, pp. 409–418, Nov. 2012.

[4] S. Okamoto and Y. Yamada, "An objective index that substitutes for subjective quality of vibrotactile material-like textures," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ Int. Conf. on*, Sept 2011, pp. 3060–3067.

[5] S. Bensmaïa, M. Hollins, and J. Yau, "Vibrotactile intensity and frequency information in the pacinian system: a psychophysical model," *Perception & Psychophysics*, vol. 67, no. 5, pp. 828–41, Jul. 2005.

[6] J. A. Fishel and G. E. Loeb, "Bayesian exploration for intelligent identification of textures," *Frontiers in Neurorobotics*, vol. 6, no. 4, pp. 1–20, 2012.

[7] S. Bensmaïa and M. Hollins, "Pacinian representations of fine surface texture," *Perception & Psychophysics*, vol. 67, no. 5, pp. 842–854, Jul. 2005.

[8] I. Hwang and S. Choi, "Perceptual space and adjective rating of sinusoidal vibrations perceived via mobile device," in *Proceedings of the 2010 IEEE Haptics Symposium*, Washington, DC, USA, 2010, HAPTIC '10, pp. 1–8.

[9] G. Park and S. Choi, "Perceptual space of amplitude-modulated vibrotactile stimuli," in *World Haptics Conference, IEEE*, June 2011, pp. 59–64.

[10] R. T. Verrillo and G. A. Gescheider, "Enhancement and summation in the perception of two successive vibrotactile stimuli," *Perception & Psychophysics*, vol. 8, no. 2, pp. 128–136, Mar. 1975.

[11] A. L. Kaas, M. C. Stoeckel, and R. Goebel, *The neural bases of haptic working memory*, In Human Haptic Perception: Basics and Applications, Springer Verlag, 2008.

[12] N.A. Macmillan and C.D. Creelman, *Detection theory: A user's guide (2nd Ed.)*, Mahwah, NJ: Erlbaum, 2005.

[13] I. Hwang, J. Seo, M. Kim, and S. Choi, "Vibrotactile perceived intensity for mobile devices as a function of direction, amplitude, and frequency," *IEEE Transactions on Haptics*, vol. 6, no. 3, pp. 352–362, 2013.