IEEE International Conference on Systems, Man, and Cybernetics (SMC), Oahu, Hawaii, USA, October 2023.

Audio-based Roughness Sensing and Tactile Feedback for Haptic Perception in Telepresence

Bastian Pätzold[1*], Andre Rochow[1], Michael Schreiber[1], Raphael Memmesheimer[1], Christian Lenz[1], Max Schwarz[1], and Sven Behnke[1]

***Abstract*— Haptic perception is highly important for immersive teleoperation of robots, especially for accomplishing manipulation tasks. We propose a low-cost haptic sensing and rendering system, which is capable of detecting and displaying surface roughness. As the robot fingertip moves across a surface of interest, two microphones capture sound coupled directly through the fingertip and through the air, respectively. A learning-based detector system analyzes the data in real time and gives roughness estimates with both high temporal resolution and low latency. Finally, an audio-based vibrational actuator displays the result to the human operator. We demonstrate the effectiveness of our system through lab experiments and our winning entry in the ANA Avatar XPRIZE competition finals, where briefly trained judges solved a roughness-based selection task even without additional vision feedback. We publish our dataset used for training and evaluation together with our trained models to enable reproducibility of results.**

Keywords: Haptics, Telepresence, Audio, Machine Learning

I. Introduction

Sensing surface properties through haptics is one of the fundamental ways, humans perceive their environment. Humans are able to perform a variety of exploratory movements with their hands and fingertips to discern aspects such as roughness, hardness, and shape of objects they manipulate [1]. It is widely understood that integrating haptics into VR, AR, and teleoperation systems is a key step towards increasing realism and acceptance of such systems [2]. Consequently, numerous methods for sensing [3], [4] and displaying [5] haptic sensations have been developed. However, these systems are often highly complex, costly, and difficult to integrate into existing teleoperation systems, especially due to size restrictions.

In this work, we present the haptic system our team NimbRo developed for the ANA Avatar XPRIZE competition[2] [6], [7]. The competition focused on intuitive and immersive telepresence in a mobile robot, including social interaction as well as manipulation capabilities. To evaluate the intuitiveness of the developed telepresence systems, briefly trained members of the judging panel had to solve through them a sequence of ten increasingly difficult tasks. The last and most difficult task focused on haptic perception, challenging the operators to discern two types of stones based on their surface roughness, i.e. "Was the Avatar able to feel the texture of the object without seeing it, and retrieve the requested one?".

[1]Autonomous Intelligent Systems, University of Bonn, Germany.
[*]Corresponding author. Email: paetzold@ais.uni-bonn.de

[2]https://www.xprize.org/prizes/avatar

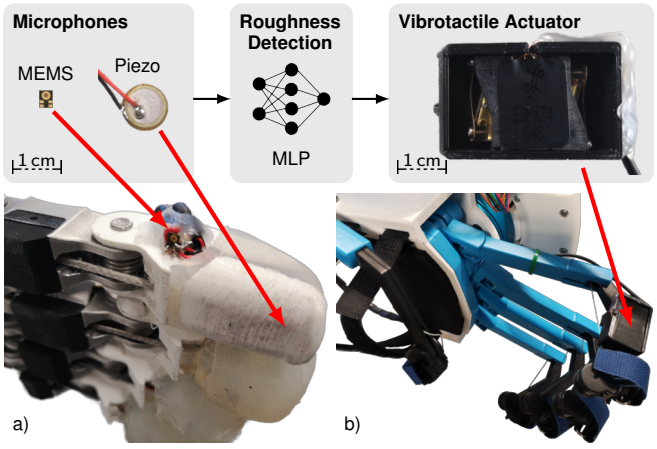

Fig. 1. Hardware implementation for roughness sensing and tactile feedback integrated with our telepresence system. Instrumented index fingers on (a) Schunk SIH robot hand and (b) SenseGlove DK1 hand exoskeleton.

In contrast to previous works, our proposed haptic sensing and display system achieves roughness sensing at very low cost by using off-the-shelf audio components. Both sensing and display components are compact and easily integrated into teleoperation systems as exemplified in Figure 1.

Our approach is based on capturing audio signals using two different types of microphones. These audio signals are analyzed by a neural network trained on a custom dataset providing exemplary surface contacts with various stones and other objects. The operator is notified about the presence and roughness of the perceived surface through low-latency vibratory feedback, which aims to convey an intuitive sense of touch that does not require special training of the operator.

The system was successfully evaluated at the ANA Avatar XPRIZE finals, where three different operator judges solved the haptic perception task as well as all other tasks in the fastest time, winning our team NimbRo the $5M grand prize.

In summary, our contributions include:

1) a compact and low-cost hardware design of both sensing and display components,
2) a learning-based method for online, low-latency and high temporal resolution roughness analysis, and
3) an evaluation of this system in the competition as well as in fully reproducible offline experiments.

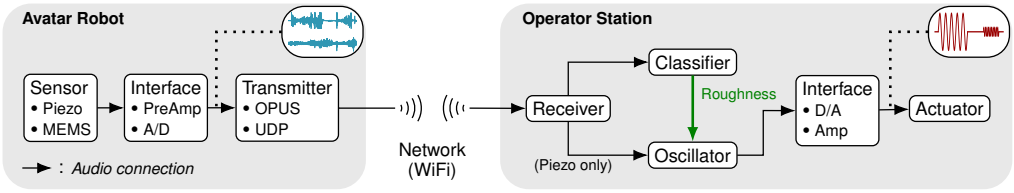

Fig. 2. Proposed end-to-end pipeline for audio-based roughness sensing and tactile feedback in telepresence applications.

II. Related Work

A. Tactile Sensing

Tactile sensors are based on a wide range of sensing principles, including capacitance, resistance, pressure, magnetism, and optics. For example, Fishel and Loeb [3] introduced the *BioTac* tactile sensor, based on an incompressible liquid as an acoustic conductor. In addition to its capability of measuring shear forces, skin stretch and temperature, it detects vibrations with up to $1040\,\mathrm{Hz}$ using a pressure sensor. By using only a single sensor per fingertip, it has a low spatial resolution, though. *GelSight*, proposed by Yuan *et al.* [8], is capable of measuring high-resolution geometry as well as local and shear forces by visually observing the deformation of an elastomer sensor surface with an embedded camera.

Despite the promising capabilities of such devices, they suffer from two disadvantages from our point of view: First, their size might be considered too large for integration in existing hardware solutions, or deployment in large quantities with high spatial resolution. Second, their availability and high cost limit their feasibility for numerous applications.

Our work focuses on deploying considerably smaller and lower-cost audio-based hardware. In a similar manner, Yoo *et al.* [9] describe the utilization of microphones to classify road surfaces by capturing the tire-pavement interaction noise in an automotive context. They convert the audio signals to time-frequency RGB images and feed them to a CNN. Even though the captured audio signals depend on other factors besides the road surface, such as the car speed, tire type, and wheel torque, they demonstrate the effectiveness of their approach for classifying snow and asphalt surfaces. Kurşun *et al.* [10] use a piezo acoustical sensor to analyze irregularities in materials, such as aluminum or stainless steel, occurring during manufacturing. They capture the friction sounds when moving a stylus with a diamond tip over the surface of a specimen. The controlledness of the environment allows the determination of roughness parameters [11] using classical signal processing approaches. Microphones have also been used to identify touch and swipe gestures on mobile devices [12], [13].

Most similar to our use case, Svensson *et al.* [14] help prosthetic users feel textures by stroking a microphone across object surfaces. In contrast to our system, the signal is filtered in a fixed manner, extracting the median frequency, which is then applied to the user using electrostimulation. This limits the scope to regular textures (such as mesh, rubber, etc.), where a frequency is easily extracted. In contrast, our method works on irregular surfaces such as natural stones.

B. Tactile Rendering

Tactile actuators are integrated into numerous devices, such as phones or game controllers. Due to the rising interest in telepresence and VR applications, a wide range of haptic displays – typically referred to as *Haptic Gloves* – emerged in recent years [15], [16], promising to convey realistic kinesthetic and tactile feedback. Typically, tactile feedback is achieved by displaying vibrations utilizing eccentric rotating masses, linear resonant actuators, or piezoelectric actuators. Their limitations often include support for only a single resonant frequency, poor intensity resolution, and slow response times. Instead, we utilize an acoustic actuator to address these issues while maintaining comparable size and costs.

III. Method

In this section, we describe our method for sensing and actuation in detail. Figure 2 gives an overview of the pipeline.

A. Sensing

The surface point of the avatar robot to be provided with roughness sensing capabilities (e.g. the tip of an index finger) is equipped with two microphones (Fig. 3). A piezo microphone is attached directly to the inside of the chosen surface to measure vibrations within the robot structure, while a MEMS-type microphone is placed in close proximity ($\sim 2\,\mathrm{cm}$) to the outside of the chosen surface measuring vibrations in the air around it. Once the surface makes contact with and slides over the unknown texture of an object, a sound gets induced into both microphones. These audio signals are then leveraged to classify the unknown texture as either *smooth* or *rough*.

B. Classical Detection

Initially, we attempted to find a direct mapping between the piezo microphone and our haptic display, utilizing classical approaches like dynamics processors and filters. While it seemed easy for our team to classify the considered textures by directly hearing the piezo microphone signal, we could not find a suitable transformation to accommodate our haptic perception and the properties of the considered actuator. Our failed attempts focused on isolating frequency regions critical for this perception, enhancing their transients, and pitch-shifting them into lower registers supported by the actuator.

Next, we decoupled the classification from the signal sent to the actuator, by generating a new audio signal using a sine oscillator that conveys the haptic perception associated with the classification result. While we found the utilization of an oscillator with varying frequency and amplitude to be

intuitive and convincing for conveying different textures, the approach to classification was not satisfactory for our application. In general, we found that rough textures induce louder and more transient signals into the piezo microphone than smooth textures, but the ambiguous amount of pressure and speed applied by the operator mask these effects. Likewise, a hand-held solid stone induces a signal that differs strongly in level and frequency spectrum from a hollow stone mounted inside a box, although their textures are very similar.

In summary, while we found a functioning configuration for a limited number of objects and scenarios, the classical approach lacked generalization across situations and users.

C. Learned Detection

Instead of hand-designed filters, we opted for a learning-based approach. As the teleoperation task demands low-latency haptic feedback, we update the prediction with every received audio buffer ($\sim 10\,\mathrm{ms}$) by constructing chunks that have access to $256\,\mathrm{ms}$ of the past. After low-pass filtering and reducing the sampling frequency, we calculate the FFT and concatenate the norm of both signals. Experiments showed a sampling frequency of $2\,\mathrm{kHz}$ to be sufficient for representing the relevant features for the described task. Classification is then performed by an MLP with 15 hidden layers of which ten layers, with 256 hidden units each, are equipped with residual connections for a better gradient flow. Experiments have shown that the classification accuracy is increased when the unnormalized input of both microphones is used, maintaining the relative loudness differences between them. When sliding over smooth surfaces – in contrast to rough ones – the MEMS microphone's level tends to be significantly quieter than that of the piezo microphone. Such patterns can easily be learned by a neural network and should therefore not be discarded through normalization. In fact, they constitute our motivation for deploying the additional MEMS microphone.

During inference, we detect the loudness of the piezo microphone in real time and compare it against a preset threshold that slightly exceeds its noise floor. This allows to distinguish between *contact* and *no contact* situations, which is used to gate the classification output of the network.

D. Actuation

The classification results are used to update the amplitude and frequency of a simple sine oscillator. In the case of a *smooth* result, we set it to a low amplitude and a high frequency (e.g. $120\,\mathrm{Hz}$), while for a *rough* result, we set a higher amplitude and a lower frequency (e.g. $60\,\mathrm{Hz}$), aiming to convey the feel of the texture, respectively. Both parameters are low-pass filtered to produce a smooth waveform. For the *no contact* case, the amplitude is set to zero.

The generated audio signal is sent to a compact loudspeaker with a special voice-coil design, capable of reproducing frequencies perceivable by the human skin ($< 1\,\mathrm{kHz}$) [17] in the form of intense but mostly inaudible vibrations. The speaker is attached to the operator station matching the sensor position on the avatar robot (e.g. the fingertip of a hand exoskeleton).

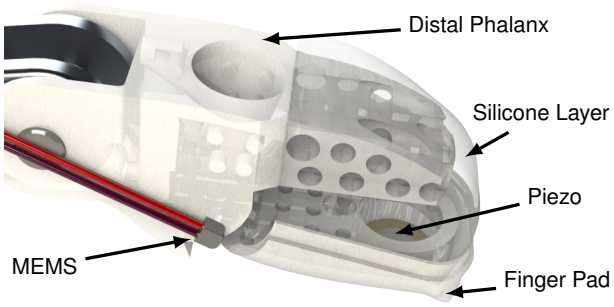

Fig. 3. CAD drawing of sensorized Schunk SIH index finger. The MEMS and piezo microphones are glued to the 3D-printed distal phalanx and finger pad. A silicone layer connects both components.

The latency of the end-to-end haptic feedback is defined by the chunk size of the classification network, the buffer size of the avatar-side and operator-side audio systems, as well as the network transmission latency between both systems.

The haptic feedback allows the operator to intuitively distinguish between the situations: *no contact*, *contact with a smooth texture*, and *contact with a rough texture*. The operator's haptic perception of smooth textures can be described as *fizzy*, while the perception of rough textures might be described as *bumpy*. The high temporal resolution of the ternary classifier allows the operator to estimate the degree of roughness and to identify local irregularities.

IV. Implementation

A. Avatar Robot

Both microphones are attached to the left index finger of our avatar robot's *Schunk SIH* hand (Fig. 1). We replaced the original index finger with a 3D-printed distal phalanx and finger pad (Fig. 3). Both feature holes which allow a silicone layer to connect them. The piezo microphone is glued to the inside surface of the finger pad. The MEMS microphone is placed on the side of the finger, where it is close to the fingertip, but does not interfere during manipulation tasks. To avoid the proximity effect affecting the network's classification performance, we chose a MEMS microphone that is omnidirectional and thus does not exhibit proximity. The silicone layer is slightly compliant and decouples the finger pad from the rest of the robot, preventing vibrations to spill over into the piezo microphone. The finger pad shape is designed to allow for sliding over a wide range of textures without getting stuck, while also producing a suitable amount of vibrations in the finger to allow for reliable classification.

The microphones are connected to a *Focusrite Scarlett 2i2* interface, which is used for pre-amplification and A/D conversion. We set both inputs to *high impedance* mode, which maximizes the microphones' frequency response. The digital audio signals are forwarded and processed using the *JACK Audio Connection Kit* which operates on top of the *Advanced Linux Sound Architecture*. Both signals are transmitted from the avatar robot to the operator station by a low-latency UDP transmitter utilizing the *OPUS audio codec* [18].

B. Operator Station

The operator station receives the audio signals of both microphones within a similar JACK environment as described for the avatar robot. Both signals are forwarded to the classification network as described in Section III-A. The confidence of the classifier modulates the frequency and amplitude of a sine oscillator. A *rough* classification targets a frequency of $60\,\mathrm{Hz}$ and a level of $0\,\mathrm{dBFS}$. Correspondingly, a *smooth* classification targets a frequency of $120\,\mathrm{Hz}$ and a level of $-25\,\mathrm{dBFS}$. Finally, we measure the loudness of the signal originating from the piezo microphone to discern *contact* and *no contact* situations. The low noise floor and mechanical decoupling of the piezo microphone allow us to easily find a suitable fixed threshold parameter to facilitate a reliable and sensitive way of measuring contact. A *no contact* situation then overwrites the roughness classifier and modulates the amplitude of the oscillator to a target level of $-\infty\,\mathrm{dBFS}$. We low-pass filter amplitude and frequency modulations to prevent artifacts in the generated waveform arising from altering classification results.

The generated audio signal is forwarded to our vibrotactile actuator shown in Fig. 1. It has been extracted from a *Lofelt Basslet*, a wearable consumer device designed to provide the sensation of bass when listening to music through headphones. It offers fast acceleration response across its frequency range of $35\,\mathrm{Hz}$ to $1\,\mathrm{kHz}$. Originally, the device is designed to be wrist-worn, houses a rechargeable battery, and offers audio connectivity via Bluetooth. Instead, we extract its vibrotactile actuator, embed it in a small-footprint 3D-printed case and drive it using the mainboard's onboard soundcard for D/A conversion and a *Fosi Audio TP-02* subwoofer amplifier. The case is attached to the left index finger of the *SenseGlove DK1* hand exoskeleton worn by the operator, inducing vibrations from above the tip.

As the chunk size processed by the MLP matches the buffer size of 512 samples set on both, the avatar robot's and the operator station's audio systems running with a sampling frequency of $48\,\mathrm{kHz}$, the latency of the entire audio system is $21\,\mathrm{ms}$ (omitting further network transmission delays).

C. Data Acquisition and Network Training

1) Dataset: We recorded a custom dataset using our instrumented robot hand, making contact with and sliding over various textures with multiple patterns and intensities. It consists of a training set with 20 objects (Fig. 4), and a test set with two objects (Fig. 6). Each object is manually labeled as either *rough* or *smooth*. We deliberately chose objects where this distinction is explicit. Since the task description in the Avatar XPRIZE competition finals clearly defined the requirement to classify the texture of artificial stones, we include various artificial stones in the dataset. However, to improve generalization, we also include natural stones and other textures such as ingrain wallpaper and a wooden table surface. Some objects are measured multiple times under varying acoustical scenarios (handheld, on a table, inside a box, etc.) to improve domain robustness. For each object, we obtain seven recordings with a duration of $30\,\mathrm{sec}$, including

Fig. 4. Dataset objects for classification of rough and smooth textures. Rough object labels are indicated by green squares.

light, medium, and strong pressure levels, with long and short strokes each, as well as a recording where we apply longer continuous wiggles, to support other interactive sensing approaches w.r.t. the operator. As inference is performed on the operator station, we encode the training data using the *OPUS audio codec*, mimicking transmission effects.

2) Training: We split the training data into chunks of $256\,\mathrm{ms}$ and adopt the label of the respective file if the RMS loudness of the chunk exceeds a threshold determining a valid contact, similar to the threshold used to mute the oscillator output. This ensures that all labeled chunks correspond to surface contacts, but conversely, not every contact is assigned to a labeled chunk. Without consideration of these unlabeled chunks, the network would show unpredictable behavior at inference time. Therefore, we introduce a third *non-valid* class (not utilized during inference) which is comprised of chunks below the set threshold. As the specific value of this threshold varies between experiments we report it in the evaluation.

The network is trained for five epochs with a batch size of 6000 chunks. We use a negative log-likelihood loss and the Adam optimizer with a learning rate of $1e\text{-}4$. We add Gaussian noise to each chunk to prevent overfitting and improve generalizability. This is particularly important, as external noises captured by the MEMS microphone or the audio circuitry might induce disturbances.

Both our dataset and the trained models are made public to enable reproducibility of results[3].

[3]https://github.com/AIS-Bonn/Roughness_Sensing

TABLE I

GENERAL MODEL CONFUSION MATRICES.

Competition runs		Response	
		Rough	Smooth
Stone	Rough	11.3 %	47.2 %
	Smooth	0.9 %	40.5 %

Test set (Fig. 6)		Response	
		Rough	Smooth
Stone	Rough	24.0 %	26.5 %
	Smooth	8.1 %	41.3 %

The system is tuned to produce a low false-positive rate (i.e. smooth surfaces classified as rough). Results are obtained using the general model variant.

TABLE II

ACCURACY OF MODEL VARIANTS.

	Competition runs					
	Rough		Smooth		Test set	
Model	Day 1	Day 2	Day 1	Day 2	Rough	Smooth
General	**0.264**	0.141	**1.000**	0.929	**0.476**	0.836
Fine-tuned	0.239	**0.190**	**1.000**	**0.959**	0.459	**0.999**
Piezo-only	0.482	0.117	0.998	0.692	0.630	0.736

We show accuracies for each class, split over the competition days and model variants. Accuracies reflect the ratio of correctly identified chunks.

V. EVALUATION

Our system has been evaluated in several steps, focusing on quantitative analysis of model training, as well as intuitiveness and immersion in a longer integrated mission.

A. Quantitative Analysis

We compare two model variants that differ in the data used during training. First, we propose a general variant trained using the entire dataset and the threshold for distinguishing contact set to $-26\,\mathrm{dBFS}$, slightly exceeding the noise floor of the piezo microphone. Table I shows the confusion matrices of the general model variant for both the test set and the competition runs. Please note that we explicitly tuned the model to produce low false-positive rates. Due to the vibration motor inside the actuator being slow in its response relative to the prediction rate, and the vibration intensity of rough classification results set relatively high, even misclassifications of single chunks can give the operator the false impression of sensing a rough surface. On the other hand, the correct classification of only several chunks suffices to convey the desired impression when sliding the finger over a rough texture.

Second, we evaluate a fine-tuned model variant optimized for participation in the competition, using a reduced set of training objects including samples of the stones encountered during the competition runs (Fig. 6). Here, we set the threshold for distinguishing contact dynamically to $50\,\%$ of the RMS loudness of all files with the respective label. This increases the amount of *non-valid* classification results when applying light pressure onto an object at the benefit of further decreasing the false-positive rate. Table II compares the classification accuracy of both model variants during the competition runs and for the test set. While the accuracy of classifying smooth objects is very high for both model variants, the fine-tuned variant substantially outperforms the general one here, but falls slightly short w.r.t. rough objects.

Furthermore, to justify the usage of the additional MEMS microphone, we show the accuracy of the general model variant using only the piezo microphone during training and inference. The low accuracy for classifying smooth textures and the associated false-positive rates of $3.3\,\%$ for the competition runs and $15.5\,\%$ for the test set are insufficient for our application.

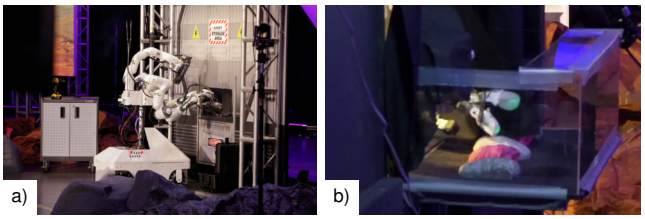

Fig. 5. Our avatar robot during the roughness sensing and stone retrieval task. a) Approaching and reaching through the box opening covered by a curtain. b) Sensing one of the stones inside the box with the instrumented finger.

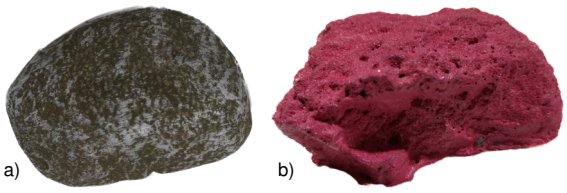

Fig. 6. Samples of the (a) smooth and (b) rough stone textures encountered during the roughness sensing and stone retrieval task in the competition.

B. ANA Avatar XPRIZE Competition

Fig. 5 shows our avatar robot during the last task in the finals testing event of the ANA Avatar XPRIZE competition. During this task, the operator was required to find and retrieve one of the rough stones, purely based on their haptic perception. In particular, there were five stones lined up on an anti-slip mat in a small box, with an opening that blocked the operator's vision through a curtain. Three of the stones had a smooth texture, while two had a rough texture and were highlighted in pink color, which was only relevant for the audience to distinguish the stones. Fig. 6 shows a close-up of sample textures encountered in the competition.

In total, the task was encountered up to three times per team during the event. Each time, a different operator judge was controlling our avatar robot. The operators were members of the XPRIZE jury and impartial in their judgment of task completion. They were trained for $45\,\mathrm{min}$, directly before the run, to familiarize themselves with the system. However, only a fraction of this time was allocated to training for this specific task, as nine previous tasks needed to be completed to advance to the final task. All three task attempts were successful. Fig. 7 and Fig. 8 show the measured audio and generated feedback signals of Days 1 & 2, respectively.

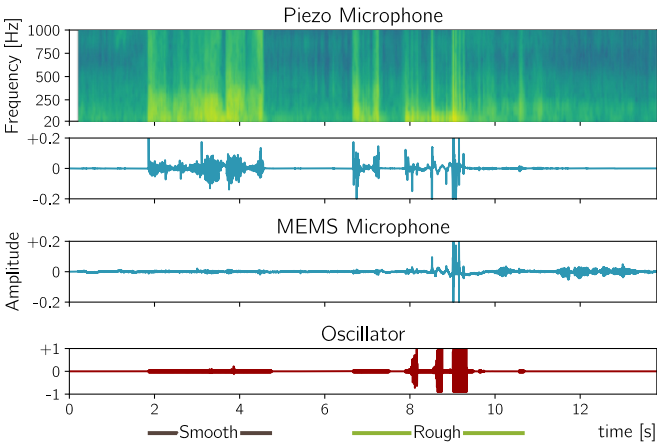

Fig. 7. Microphone and oscillator signals during the roughness sensing and stone retrieval task on Day 1. Ground truth times of contact and rock type are shown at the bottom.

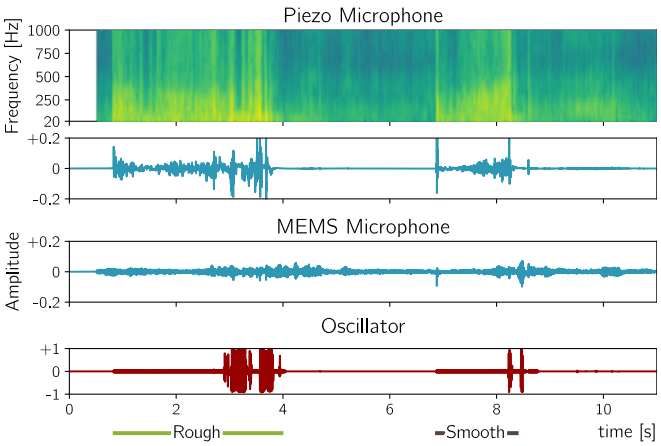

Fig. 8. Microphone and oscillator signals during the roughness sensing and stone retrieval task on Day 2. Ground truth times of contact and rock type are shown at the bottom. Note the false positive classification at the end of the smooth stone, where the finger slipped off and hit the surface underneath. The operator correctly interpreted this as the edge of the stone.

While the first run on Qualification Day was not public and results are not available, the runs on Testing Days 1 & 2 were broadcasted by the organizers[4], allowing for a comparison with all other teams that completed the task. Table III shows that we completed the task considerably faster than any other team. However, it is worth mentioning that other factors besides the haptic feedback may have contributed to the reported times.

VI. Conclusion

We presented an audio-based haptic teleperception system that is based on low-cost, compact components. The system was proven to be very effective at the ANA Avatar XPRIZE competition finals, winning the first prize. Even though the system was mostly trained and tested on stone surfaces, the method can be adapted easily to other surface kinds by collecting the appropriate training data.

[4]https://www.youtube.com/watch?v=EmESa2Olq4c

TABLE III

Task completion times.

	NimbRo	Pollen Robotics	Northeastern	Avatrina
Day 1	**1:06**	2:24	N/A	4:48
Day 2	**1:02**	1:59	9:27	N/A

Time is given in min:sec and includes roughness sensing and stone retrieval. N/A: not attempted.

References

[1] S. J. Lederman and R. L. Klatzky, "Extracting object properties through haptic exploration," *Acta Psychologica*, vol. 84, no. 1, pp. 29–40, 1993.

[2] C. Wee, K. M. Yap, and W. N. Lim, "Haptic interfaces for virtual reality: Challenges and research directions," *IEEE Access*, vol. 9, pp. 112 145–112 162, 2021.

[3] J. A. Fishel and G. E. Loeb, "Sensing tactile microvibrations with the BioTac - Comparison with human sensitivity," *International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, pp. 1122–1127, 2012.

[4] A. C. Abad and A. Ranasinghe, "Visuotactile sensors with emphasis on GelSight sensor: A review," *Sensors*, vol. 20, no. 14, pp. 7628–7638, 2020.

[5] A. Adilkhanov, M. Rubagotti, and Z. Kappassov, "Haptic devices: Wearability-based taxonomy and literature review," *IEEE Access*, vol. 10, pp. 91 923–91 947, 2022.

[6] M. Schwarz, C. Lenz, A. Rochow, M. Schreiber, and S. Behnke, "NimbRo Avatar: Interactive immersive telepresence with force-feedback telemanipulation," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 5312–5319, 2021.

[7] M. Schwarz, C. Lenz, R. Memmesheimer, B. Pätzold, A. Rochow, M. Schreiber, and S. Behnke, *Robust immersive telepresence and mobile telemanipulation: NimbRo wins ANA Avatar XPRIZE finals*, 2023. arXiv: 2303.03297.

[8] W. Yuan, S. Dong, and E. H. Adelson, "Gelsight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, p. 2762, 2017.

[9] J. Yoo, C.-H. Lee, H.-M. Jea, S.-K. Lee, Y. Yoon, J. Lee, K. Yum, and S.-U. Hwang, "Classification of road surfaces based on CNN architecture and tire acoustical signals," *Applied Sciences*, vol. 12, no. 19, 2022.

[10] K. Kurşun, F. Güven, and H. Ersoy, "Utilizing piezo acoustic sensors for the identification of surface roughness and textures," *Sensors*, vol. 22, no. 12, 2022.

[11] ISO 21920-2:2021, "Geometrical product specifications (GPS) — Surface texture: Profile — Part 2: Terms, definitions and surface texture parameters," p. 78, 2021.

[12] K. Sun, T. Zhao, W. Wang, and L. Xie, "VSkin: Sensing touch gestures on surfaces of mobile devices using acoustic signals," *ACM International Conference on Mobile Computing and Networking*, pp. 591–605, 2018.

[13] P. Lopes, R. Jota, and J. A. Jorge, "Augmenting touch interaction through acoustic sensing," *ACM International Conference on Interactive Tabletops and Surfaces*, pp. 53–56, 2011.

[14] P. Svensson, C. Antfolk, A. Björkman, and N. Malešević, "Electrotactile feedback for the discrimination of different surface textures using a microphone," *Sensors*, vol. 21, no. 10, p. 3384, 2021.

[15] M. Caeiro-Rodríguez, I. Otero-González, F. A. Mikic-Fonte, and M. Llamas-Nistal, "A systematic review of commercial smart gloves: Current status and applications," *Sensors*, vol. 21, no. 8, 2021.

[16] J. Perret and E. Vander Poorten, "Touching virtual reality: A review of haptic gloves," *International Conference on New Actuators*, pp. 1–5, 2018.

[17] J. Bolanowski S. J., G. A. Gescheider, R. T. Verrillo, and C. M. Checkosky, "Four channels mediate the mechanical aspects of touch," *Journal of the Acoustical Society of America*, vol. 84, no. 5, pp. 1680–1694, 1988.

[18] J. Valin, K. Vos, and T. Terriberry, "Definition of the Opus audio codec," *Internet Requests for Comments*, RFC 6716, 2012.