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

Bastian Pätzold, Andre Rochow, Michael Schreiber, Raphael Memmesheimer,
Christian Lenz, Max Schwarz, and Sven Behnke
Autonomous Intelligent Systems
University of Bonn, Germany
paetzold@ais.uni-bonn.de

***Abstract*—Haptic perception is incredibly 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 haptic actuator displays the result to the human operator. We demonstrate the effectiveness of our system through experiments and our winning entry in the ANA Avatar XPRIZE competition finals, where impartial 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.**

***Index Terms*—Haptics, Audio, Telepresence, Machine Learning**

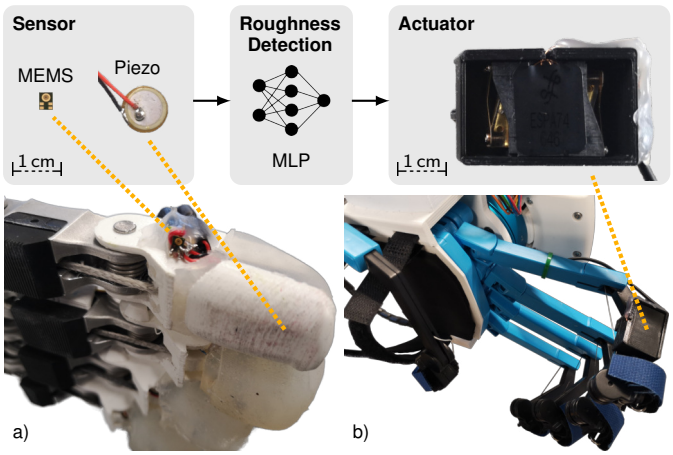

Fig. 1. Hardware implementation for roughness sensing and tactile feedback integrated with our telepresence system. a) Instrumented index finger on Schunk SVH hand. b) Instrumented index finger on SenseGlove DK1 hand exoskeleton.

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.

These systems are often very complex, though, and require too much space, especially on the side, where, if we aim at using humanoid-like robot hands, the volume of a fingertip has to suffice.

In this work, we showcase the haptic system our team NimbRo developed for the ANA Avatar XPRIZE competition[1] [6], [7]. The competition focused on immersive telepresence in a mobile robot including social interaction as well as manipulation capabilities. To evaluate the intuitiveness of the developed telepresence systems, judges had to solve a sequence of increasingly difficult tasks after very little training (30 min). The most difficult task to be performed in the finals was to discern two stone types based on their surface roughness.

In contrast to previous works, our 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 of very small size and are easily integrated into teleoperation systems. The audio signal captured by two microphones is analyzed by a neural network, which was trained on a dataset of exemplary surfaces. The operator is notified about the presence and roughness of the perceived surface through low-latency vibratory feedback.

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

In summary, our contributions include:

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

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

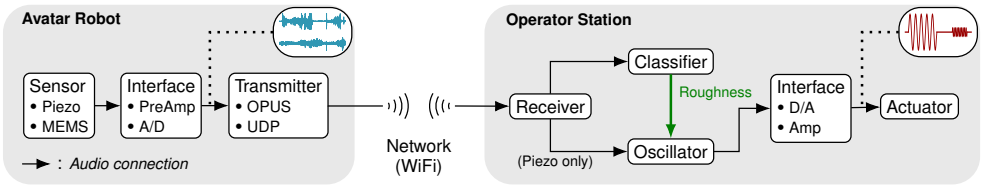

Fig. 2. Proposed end-to-end pipeline for audio-based roughness sensing and tactile feedback in telepresence applications.

II. Related Work

A. Tactile Sensing

Tactile sensors are based on a wide range of sensing principles, including capacitance, resistance, 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/shear forces by visually observing the deformation of an elastomer sensor surface.

Despite the promising capabilities of such devices, they suffer from two drawbacks: First, their size is too large for deployment in large quantities with high spatial resolution, or integration in existing hardware solutions. Second, their high cost limits their feasibility for many 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. They convert the audio signals to time-frequency RGB images and feed them to a CNN. Even though the captured audio signals also depend on other factors than 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 the specimen. This controlled environment allows the determination of roughness parameters [11] using classical signal processing approaches. Microphones have also been used to identify touch and swiping gestures in smartphone contexts [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 utilizing eccentric rotating masses, linear resonant actuators or piezoelectric actuators to display vibrations. Their limitations often include supporting only a single resonant frequency, poor intensity resolution, as well as slow response times. Instead, we utilize an acoustic actuator to address these issues while maintaining comparable size and costs.

III. Method

We now describe our method in detail. Figure 2 gives an overview of the data flow.

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

Next, we tried to first classify the piezo microphone signal using similar DSP techniques, to then generate a new audio signal using a sine oscillator that conveys the haptic perception associated with the classification result. In general, rough textures induce louder signals into the piezo microphone than

smooth textures. However, the ambiguous amount of pressure applied by the operator masks this effect. Likewise, a hand-held solid natural stone induces a signal that differs strongly in level and frequency spectrum from a hollow artificial 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 the piezo and MEMS microphones are used. 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 a MEMS microphone in addition to the piezo 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 distinguishing between *contact* and *no contact* situations, which is used to either pass or block the network classification output.

D. Actuation

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

The audio signal generated by the oscillator is fed into a compact loudspeaker with a special voice-coil design, capable of reproducing frequencies between $10\,\mathrm{Hz}$ and $250\,\mathrm{Hz}$ 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. a fingertip of a hand exoskeleton).

The latency of the end-to-end haptic feedback is defined by the chunk size of the network, the buffer size of the avatar- and operator audio systems, as well as the transmission latency between them. The haptic feedback allows the operator to intuitively distinguish between *no contact*, *contact with a smooth texture* and *contact with a rough texture* situations. The operator's haptic perception of smooth textures can be described as *fizzy*, while the perception of rough textures might be described as *bumpy*.

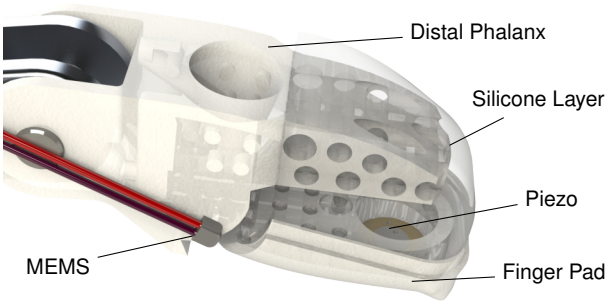

Fig. 3. CAD drawing of sensorized Schunk SIH index finger. The Piezo and MEMS microphone are glued to the 3D printed distal phalanx and finger pad. A silicone layer connects both components.

IV. Implementation

A. Avatar Robot

Both microphones are attached to the left index finger of the avatar robot's *Schunk SIH* hand (Fig. 1). We replaced the original index finger with two 3D printed components: The distal phalanx and the finger pad (Fig. 3). Both feature hollow cylinders inside which allow a silicone layer to connect both components. 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. In order to avoid any implications caused by the proximity effect on the network's classification performance, we choose this microphone to have an omnidirectional polar pattern. 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* [17].

B. Operator Station

The operator station receives the audio signals of both microphones within a similar JACK environment as described for the avatar robot. First, the signal originating from the piezo microphone is routed directly to the headphones of the *Valve Index* head-mounted display that the operator is wearing.

This helps to convey a more detailed and complete haptic perception of the texture in question.

Then, both audio signals are fed to the classification network as described in Section III-A. The confidence of the *rough* class is used to modulate the amplitude of a sine oscillator set to a frequency of $60\,\mathrm{Hz}$ and a maximum level of $0\,\mathrm{dBFS}$. We apply a low-pass filter on this modulation to prevent artifacts in the generated waveform arising from immediate jumps in confidence. A second sine oscillator is set to a frequency of $120\,\mathrm{Hz}$ and a level of $-25\,\mathrm{dBFS}$. Before mixing the two generated signals, we use a side-chain gate to mute the latter oscillator in *no contact* situations. This is achieved by feeding the audio signal originating from the piezo microphone to the respective side-chain input. The low noise floor and mechanical decoupling of the piezo microphone allow us to easily find a suitable threshold parameter to be set within the sidechain gate, in order to facilitate a reliable and sensitive way of sensing contact. Again, nonzero *attack* and *release* parameters prevent the gate from inducing artifacts arising from fast amplitude modulations.

The generated audio signal is forwarded to our haptic display shown in Fig. 1. Its actuator is extracted from a *Lofelt Basslet*, which is a wearable consumer device designed to convey the sense of bass while listening to music via headphones. Originally, the device is wrist-worn, houses a rechargeable battery and offers audio connectivity via Bluetooth. Instead, we embed its haptic actuator in a small-footprint 3D printed case and drive it using the mainboard-integrated 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.

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 system 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 labeled objects, as shown in Fig. 4, and two test objects Fig. 6. Both sets feature an equal number of *rough* and *smooth* textured objects. As the task description in the Avatar XPRIZE finals was clear in defining the requirement of classifying the texture of stones, we include various natural and artificial stones in the dataset. However, to improve generalization, we also included other common textures like 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 light, medium, and strong pressure levels, with long and short strokes, 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 run on the operator station, we encode all recordings using the *OPUS audio codec* before training, mimicking transmission effects.

Fig. 4. Dataset objects for classification of rough and smooth textures.

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 forward or block the oscillator output. As the specific value of this threshold varies between experiments we report it in the evaluation, respectively. 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. In the worst case, *smooth* chunks would be assigned to the *rough* class, which was particularly undesirable in the context of the competition. Therefore, we introduce a third *non-valid* class which is comprised of chunks below the set threshold.

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

Both our dataset and the trained models are made public to allow for reproducibility of results[2].

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

TABLE I
CONFUSION MATRICES

		Competition runs		Test set (see Fig. 6)	
		Response		Response	
		Rough	Smooth	Rough	Smooth
Stone	Rough	11.3 %	47.2 %	24.0 %	26.5 %
	Smooth	0.9 %	40.5 %	8.1 %	41.3 %

Notes: 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
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 different model variants.

V. EVALUATION

Our system has been developed and evaluated in several steps, focusing on a quantitative analysis of the model training as well as the 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. It is to be pointed out 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, w.r.t. 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. Accordingly, the correct classification of only several chunks suffices to convey the desired impression when sliding the finger over a rough texture.

Second, we show 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 for classifying smooth objects shows to be very high with both model variants, the fine-tuned variant does substantially exceed the general one but falls slightly shorter w.r.t. rough objects.

Furthermore, to justify the usage of the additional MEMS microphone, we show the accuracy of the general model 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.

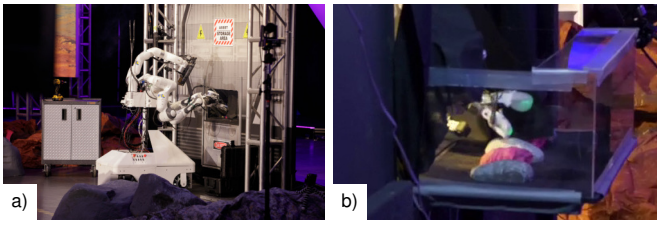

Fig. 5. Our avatar robot during the roughness sensing and stone retrieval task. a) Approaching and reaching through the box opening covered by a curtain. b) Sensing one of the stones inside the box with the instrumented finger.

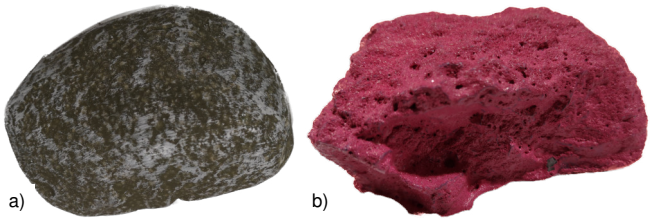

Fig. 6. Samples of the smooth (a) and rough (b) stones encountered during the competition.

B. Integrated Mission

Fig. 5 shows our avatar robot during the last task in the finals testing event of the ANA Avatar XPRIZE competition. During

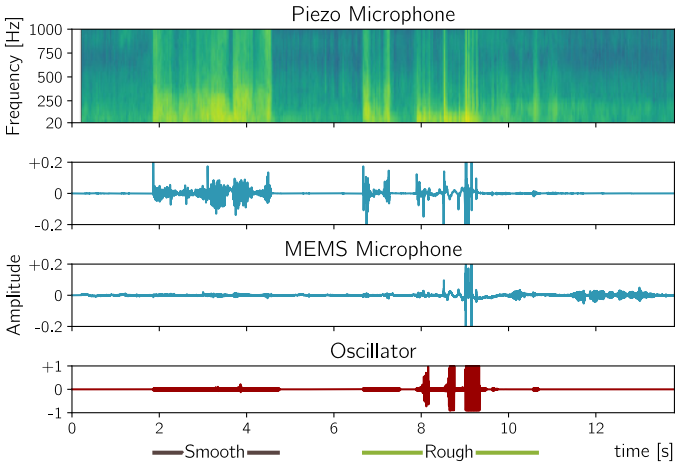

Fig. 7. Microphone input and oscillator output signals during the roughness sensing and stone retrieval task on Day 1. Ground truth times of contact and rock type are shown at the bottom.

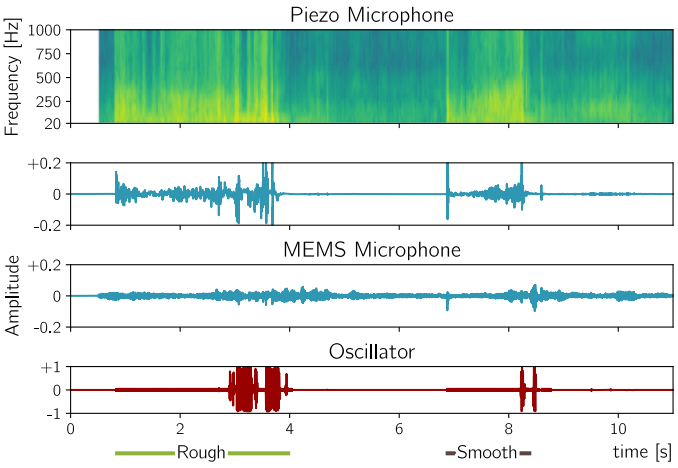

Fig. 8. Microphone input and oscillator output signals during the roughness sensing and stone retrieval task on Day 2. Ground truth times of contact and rock type are shown at the bottom. Note that the system generated a false positive (rough) detection at the end of the smooth stone, where the finger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

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 track was encountered up to three times per team during the event. Each time, a different operator 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 30 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 encounters were successful. Fig. 7 and Fig. 8 show the measured audio and generated feedback signals of Day 1 & 2, respectively.

While the first run on Qualification Day was not public and results are not available, the latter two runs on testing Days 1 & 2 were broadcasted by the organizers, 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 contributed to these times, such as fitting the hand inside the box and successfully grasping the stone after identification.

C. Operation without Sight

During our qualification run, we also demonstrated operation without direct sight, i.e. the operator was not able to see the scene and the stones through the robot's cameras as

TABLE III
TASK COMPLETION TIMES.

	NimbRo	Pollen Rob.	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.

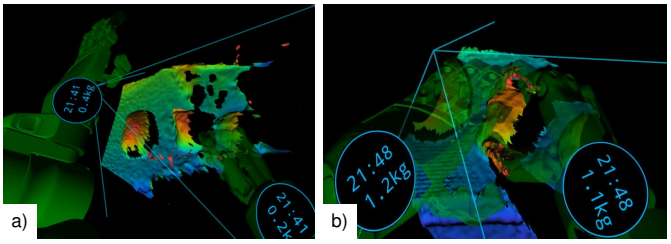

Fig. 9. VR visualization for operation without sight. The operator views the scene geometry stereoscopically in VR, with height encoded as color (blue to red) and robot arms/hands as a green overlay. a) Using a thumb gesture, the operator activates the visualization and locates the stones in the live geometry view. b) After fixating a particular stone with the right hand, the operator disables updates to the geometry. The left hand (carrying the camera) is now free to move, leaving the camera frustum (cyan) fixed in place. The operator moves the index finger equipped with our sensor assembly over the stone to feel the roughness. We show actual screenshots from our qualification run.

usual. In this scenario, it is difficult for the operator to even find the stones in order to touch them. For this reason, we developed a 3D visualization based on geometry captured by a depth camera mounted in the left palm of the robot. It allows the operator to locate the stones, hold them in place using the right hand, and move the instrumented index finger of the left hand over them. We note that the system is designed to not give away the surface roughness: The measured depth data is adaptively downsampled to remove surface details, leaving only the rough geometric shape. Using this depth-based guidance, the operator judge was able to solve the task without live color view, which demonstrates that our haptic feedback alone is sufficient for the classification of the stones.

VI. CONCLUSION

We demonstrated a haptic teleperception system consisting of a microphone sensing setup, which is low-cost and very compact, and a simple haptic rendering method on the operator side. 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.

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*, 2022.

[6] M. Schwarz, C. Lenz, A. Rochow, M. Schreiber, and S. Behnke, "NimbRo Avatar: Interactive immersive telepresence with force-feedback telemanipulation," *2021 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," *Submitted to 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2023.

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