AARHUS
UNIVERSITY
DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

AU Dataset for Visuo-Haptic Object Recognition for Robots

Lasse Emil R. Bonner[1], Daniel Daugaard Buhl[1], Kristian Kristensen[1] and Nicolás Navarro-Guerrero[1*]

[1*]Department of Electrical and Computer Engineering, Aarhus University, Finlandsgade 22, 8200 Aarhus, Denmark.

*Corresponding author's e-mail(s):
nicolas.navarro.guerrero@gmail.com

Abstract

Multimodal object recognition is still an emerging field. Thus, publicly available datasets are still rare and of small size. This dataset was developed to help fill this void and presents multimodal data for 63 objects with some visual and haptic ambiguity. The dataset contains visual, kinesthetic and tactile (audio/vibrations) data. To completely solve sensory ambiguity, sensory integration/fusion would be required. This report describes the creation and structure of the dataset. The first section explains the underlying approach used to capture the visual and haptic properties of the objects. The second section describes the technical aspects (experimental setup) needed for the collection of the data. The third section introduces the objects, while the final section describes the structure and content of the dataset.

Keywords: Multimodal Object Recognition, Data Fusion, Visuo-Haptic Object Recognition, Tactile Perception, Haptic Information, Kinesthetic Perception, Robot Perception

Downloading and Citing this dataset: Bonner, L. E. R., Buhl, D. D., Kristensen, K., & Navarro-Guerrero, N. (2021). AU Dataset for Visuo-Haptic Object Recognition for Robots. figshare. https://doi.org/10.6084/m9.figshare.14222486

1 Object Exploration

In order to create the dataset, it is necessary to capture both visual and haptic information from the selected objects. When choosing how to explore the objects, the primary consideration is to capture relevant information from both the visual and haptic modalities. Object explorations are based on the *exploratory procedures* described by Lederman and Klatzky [7, 8]. The haptic properties of the objects can be divided into two sub-categories: kinesthetic, which comprises size, shape and weight, and tactile, which comprises texture and hardness. The exploration carried out for each object can be divided into three phases:

- Visual object exploration,
- Kinesthetic object exploration,
- Tactile object exploration.

1.1 Visual Exploration

In this phase, the objects are repositioned three times to expose different faces to the camera. In addition to the three images of the object, one image of only the background is captured, which can be used to implement background subtraction when processing the data further. The three images are taken in optimal lighting conditions where artificial light from multiple angles decreases the shadows from the objects. The resolution of the images is 4640×3472. Images can later be manipulated by, for instance, scaling, applying noise and other filters to increase the difficulty of the task as needed. Despite the high quality of the images, visuo-haptic integration is still necessary to reliably classify all objects as shown in the thesis of Buhl and Bonner [4], and Kristensen [6].

1.2 Kinesthetic exploration

For Kinesthetic Exploration, "Enclosure" and "Unsupported holding" were selected [7, 8].

"Unsupported holding"

The object is lifted and held in the hand. This exploratory procedure provides information about the weight of the object. This procedure is implemented by mounting a human-sized robot hand (RH8D [3]) horizontally while placing the object in the hand. By measuring the current needed to keep the hand still, the weight of the objects can be inferred.

"Enclosure"

The object is enveloped with the hand(s) as much as possible. This exploratory procedure provides some information about the global shape of the object. This procedure is implemented by placing the object in the open RH8D hand, then the hand is closed, and the final position of the fingers is recorded.

1.3 Tactile exploration

For Tactile Exploration, inspiration from the "Lateral motion" and "Pressure" procedures was taken [7, 8].

"Lateral motion"

Sideways movements over the object help to extract information about the object's texture. This procedure is implemented by placing the object in the RH8D robotic hand, ensuring that most of the object's surface is exposed. Then another robot – NAO [2] – explores the object by moving one of its fingers across the object in a side-to-side manner. This motion creates vibrations which are recorded and later used to infer the texture of the object. This procedure will be referred to as "feel" within the dataset.

"Pressure"

Two exploratory procedures inspired by the "Pressure" procedure described by Lederman and Klatzky are implemented to capture information about the object's hardness. The "Pressure" procedure consists of applying a force to the object to infer the object's hardness. In this dataset, this procedure recorded data for two implementations. The first implementation, called "pressure-poke", uses the NAO robot to poke the object while the object is being held by the RH8D hand. The second implementation called "pressure-squeeze", is similar to the implementation of "Enclosure". The object is placed in the RH8D hand, and the hand is closed around the object. Then the fingers apply additional force in small increments. The force and number of increments are the same for all objects.

2 Experimental Setup

The experimental setup used for the data collection consists of an NAO robot, a human-sized robotic hand RH8D, a camera, an oscilloscope, contact microphones and an array of everyday objects that vary in colour, shape, size, material and weight. Clip-on contact microphones are placed in different locations on the NAO robot and the RH8D hand to measure the vibrations produced when the objects are explored. The main parts of the experimental setup are shown in Figure 1.

2.1 Camera

The camera used is an Olympus OM-D E-M10 Mark III. The camera was mounted on a tripod and at a fixed distance to the objects. The camera was set to auto, and images were taken once the object was in focus. The camera remained fixed, and only the objects were repositioned to expose different faces.

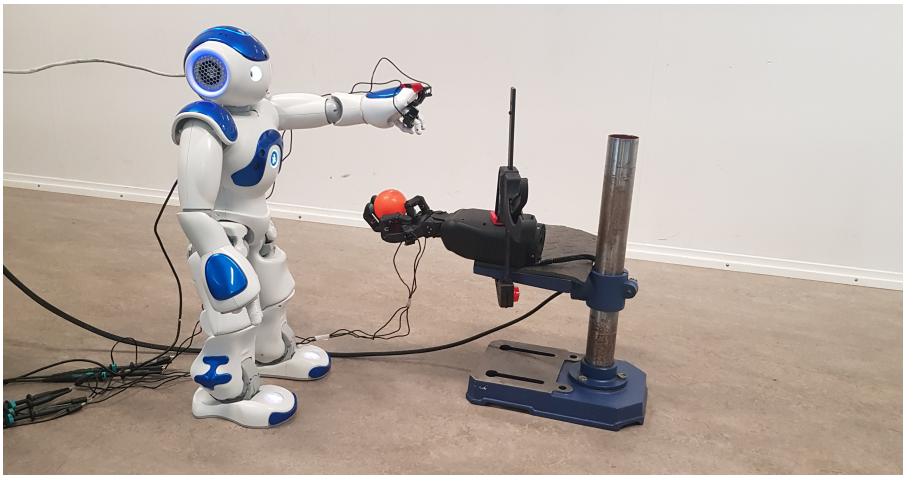

Fig. 1 Experimental setup with NAO robot, RH8D Seed Robotics hand and attached microphones.

2.2 NAO robot

The NAO robot used is version 5 with NAOqi version 2.1.14.3. The robot is used to capture the tactile properties (texture and hardness) of the objects by touching and tapping the objects as described in Section 1. The NAO robot does not have any built-in sensors that are capable of registering tactile feedback. Thus, five contact microphones are mounted in different locations on both the NAO robot and the RH8D hand. This approach is inspired by the work of Toprak et al. [9], who showed that tactile sensing is possible using contact microphones in a very similar setup.

The microphones are used to record the vibrations produced when the robot explores the objects. The robot's fingers are coated with rubber. Thus, the vibrations produced when exploring the surface of the objects are attenuated before reaching the microphones. For this reason, we created a 3D-printed thimble covered with symmetric protuberances, emulating fingerprints, placed on the robot's finger. The 3D model of the thimble can be seen in Figure 2 and can be downloaded as part of this dataset.

The thimble is printed in polylactide (PLA) and has a wall thickness of 2mm. The protuberances are 2mm high and distributed around the finger and on the tip with a distance of 0.3mm. The resolution of the thimble is not as fine as a human fingerprint due to the resolution limitations of the 3D printer used, which can print down to a precision of 0.1mm [1]. A test was conducted on a few objects from the object suite to ensure that the thimble creates more vibrations than the NAO finger itself. These experiments showed that, at a minimum, using the thimble increased the measured signal by 166%. On the most significant microphone – placed right under the object – the signal strength increased by 470% using the thimble. Despite the excellent results

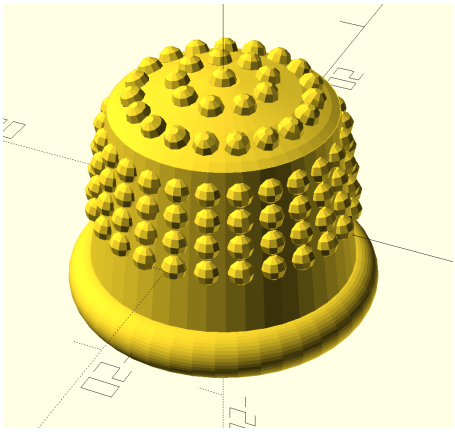

Fig. 2 The 3D model used to print the thimble.

obtained with this thimble's design, the thimble's design was not systematically optimized. Hence a thorough study of thimbles design and materials is recommended.

Although the NAO robot could also be used to record other haptic data, such as the kinesthetic properties: shape and weight, as it was done by Toprak et al. [9], the physical size of the robot poses too many constraints on the type of objects that can be used. Additionally, the NAO's hand can only be controlled with a binary signal and does not allow the implementation of the "pressure-squeeze" procedure. Thus, the RH8D hand was used for all the kinesthetic exploration procedures and the pressure-squeeze procedure instead.

2.3 RH8D Seed Robotics hand

The RH8D hand enabled us to use objects of a broader range of sizes and weight than those possible with the NAO robot. Additionally, the affordability and accessibility of the RH8D hand made it a good option to facilitate reproducibility and potentially expand the dataset. The hand is human-sized, has eight motors, and provides 19 degrees of freedom and 4096 different position readings of each motor [3]. From each actuator, real-time current readings are available and can be used to determine the weight of the objects. All these characteristics make the RH8D Seed Robotics hand a good choice to haptically explore objects.

2.4 Microphones

The microphones used are Harley Benton CM-1000 clip-on contact microphones. A PicoScope 4824 with eight channels was used to record the data from the microphones with a sampling rate of 400kHz. A consideration when choosing the sample rate is that the data should be suited for sound source localization as well as object recognition.

To perform localization, it is necessary to distinguish the time of arrival of the vibrations to each of the microphones. The speed of sound in plastic is about 2750 m/s at 20°C [5]. Because the microphones are separated by at least 2cm, it is necessary to sample with at least 275kHz to capture the time difference between them. Moreover, the NAO robot and the RH8D hand are made from different types of plastic and other materials, therefore a sample rate of 400kHz is chosen.

2.4.1 Finding the optimal microphone placement

The tactile features, texture and hardness, are reliant on well-placed microphones to capture as many vibrations created when the NAO robot interacts with the object as possible. Multiple configurations were tested to ensure optimal placement of the microphones, which was determined with a three-step procedure. The primary consideration was to increase the difference of the measurements for the different objects – e.g., a soft rubber ball and a hard wooden block.

Firstly, live measurements were carried out to test promising placements both based on the proximity to the point of contact as well as a decent contact between the microphone and the robot. The best placement candidates are listed in Table 1.

Table 1 Name and placement of microphones (channel) on the NAO and RH8D. The five best placements are shown in Figure 6.

Channel	Placement
NP	NAO palm
NF	NAO finger
NW	NAO wrist
R8RU	RH8D hand, right side, under
R8LU	RH8D hand, left side, under
R8RO	RH8D hand, right side, over
R8LO	RH8D hand, left side, over

Secondly, all the exploratory procedures were performed on a subset of objects.

Finally, the best placements were determined based on the signal strength. The placements leading to low signal strength were removed from the final setup. From the subset of objects, we focused on a soft rubber ball and a hard wooden box as these objects lead to very different signal patterns. Figure 3 and 4 show the measurements for the lateral motion procedure on a soft rubber ball and a hard wooden box, respectively.

To quantify the difference between the channels, the absolute average difference of a channel is compared across measurements on the two different objects. The differences between the channels are shown in Figure 5. The smallest difference is detected in channels NW and R8LU. Thus, these channels are not used in the final setup. The channels NP and R8RO show the most

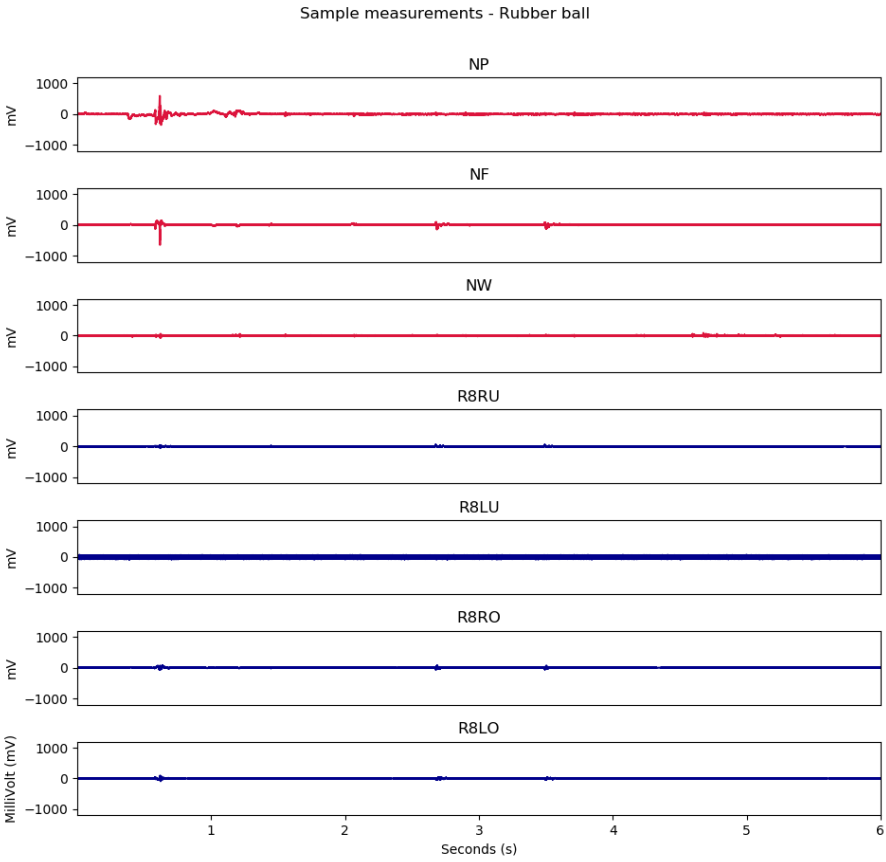

Fig. 3 Measurements from all seven microphones when exploring a soft rubber ball. Red is used for Microphones placed on the NAO, and blue is used for microphones placed on the RH8D.

significant difference and are therefore expected to provide a more significant amount of information about the haptic properties of the objects.

This leads to the final setup shown in Figure 6, consisting of five microphones from which two are on the NAO and three on the R8HD.

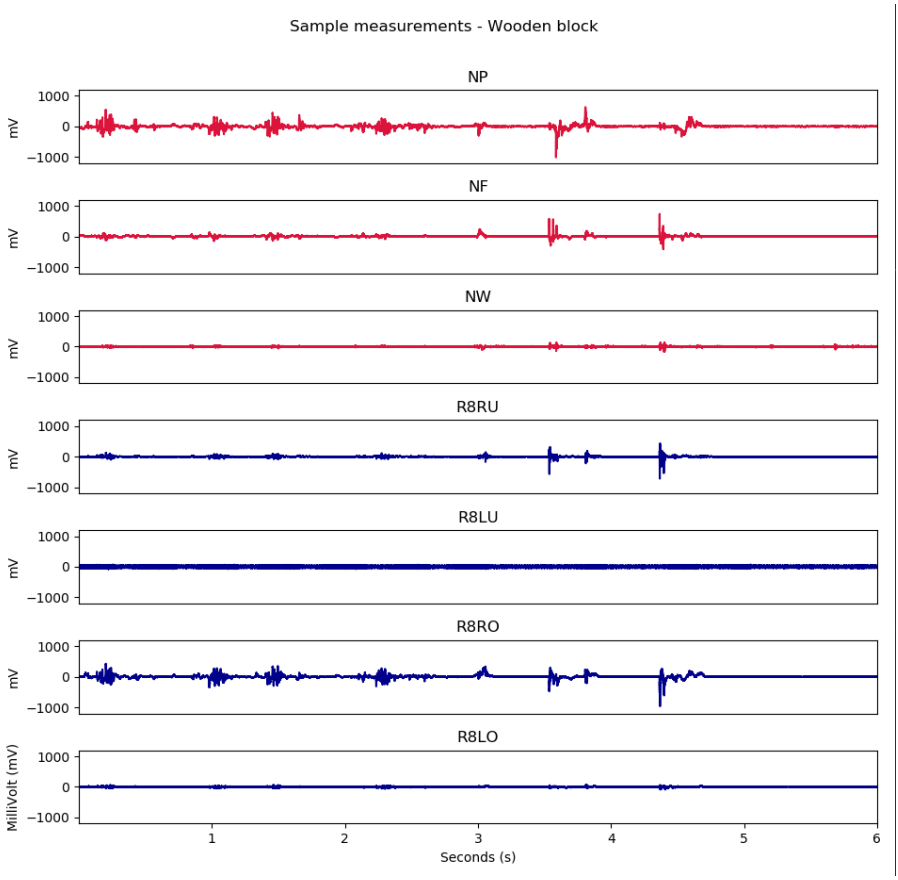

Fig. 4 Measurements from all seven microphones when exploring a hard wooden box. Red is used for Microphones placed on the NAO, and blue is used for microphones placed on the RH8D.

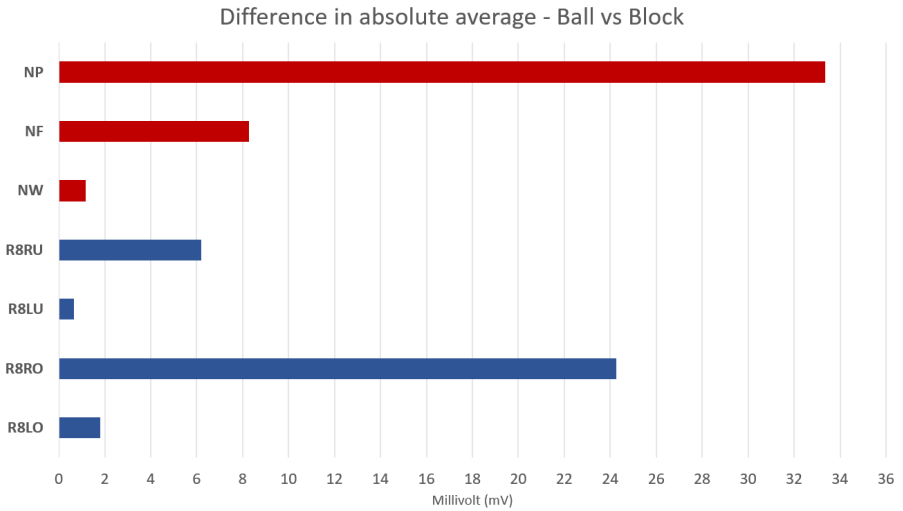

Fig. 5 Absolute average difference of the different channels while exploring a wooden box and a rubber ball. Red is used for Microphones placed on the NAO, and blue is used for microphones placed on the RH8D.

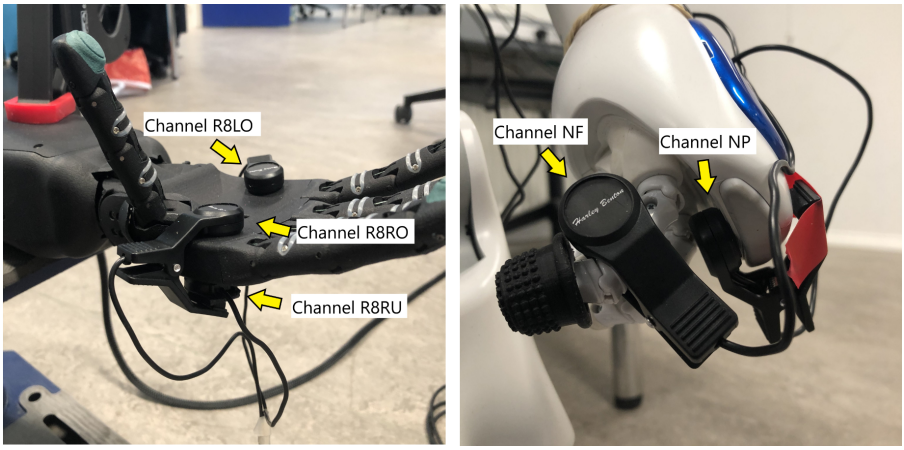

Fig. 6 Final microphone setup. Left: the R8HD robotic hand shows channels R8LO, R8RU and R8RO. Right: the hand of the NAO showing channel NF and NP.

3 Object Set

In this section, the considerations used for object selection are presented. Additionally, pictures and properties for each of the objects are listed.

3.1 Choosing the objects

The objects are chosen based on their suitability for multimodal (visuo-haptic) object recognition. We used the following criteria for the dataset:

- The dataset should contain visually ambiguous objects
- The dataset should contain haptically ambiguous objects
- The dataset should contain objects of a variety of materials and colours
- The number of objects in the dataset should exceed 50 (most visuo-haptic datasets include less than 50 objects).
- The objects should be small enough to fit in the RH8D hand but large enough to facilitate the automated exploration procedures.

Figure 7 shows an overview of the objects included in the dataset. These images are not part of the dataset but are merely used to identify the objects. The objects chosen are a combination of toys and household objects. Especially toys were found to be well suited because they come in different shapes and colours. Objects of similar shape and colour are included to create visual ambiguity. An example of this is the yellow ball and yellow lemon. Some objects are filled with different content, making them only separable by their haptic features. Examples of this are the ice cube container, which is presented as empty and filled with Play-Doh, and the velvet bag filled with 65g salt, coffee beans, and Play-Doh, respectively. Both the ice cube container and velvet bag are presented at the bottom right of Figure 7. The haptic ambiguity is included by choosing objects of similar material and shape but in a different colour. Examples of this are the balls and cubes. For further specifications of the selected objects, please see the object list uploaded with the dataset.

4 Dataset

The final dataset consists of the previously mentioned visual and haptic data. Each object is measured three times and repositioned each time. This data is collected from three sources:

- images of the objects,
- kinesthetic data from the RH8D hand,
- vibration data from the contact microphones.

The visual information consists of four images, a background image for each object and three images of different faces of the object. The distance and heading of the camera remain unchanged for all four images and across objects.

The kinesthetic data collected with the RH8D hand consists of current readings from the wrist and the positions of each of the hand's five fingers.

Fig. 7 All objects used in the dataset. Some objects are presented several times because they are filled with different substances, which can change the weight and, in some cases, texture and hardness of the object.

The current from the wrist flexion joint is measured in milliamperes (mA), and the baseline reading (without objects) is 30mA. The fingers' positions are represented by a value in degrees ranging from -180 degrees (stretched) to 180 degrees (closed). All the information is stored in CSV files. There are individual files for each exploratory procedure ("unsupported holding", "enclosure", and "pressure-squeeze") and channels for three different object repositionings. Additionally, readings from the IR proximity sensor (located at the centre of the palm) are provided. All in all, resulting in 60 files for kinesthetic data. The current data is stored in the unsupported_holding_n.csv files, and the IR proximity sensor reading is stored in the extra_n.csv files. The fingers' position data is stored in the enclosure_n.csv files. After the enclosure procedure is sampled, the object is squeezed with four different forces (400, 500, 600 and 700 units of the actuators max force in a 12bit resolution scale, i.e., from 0 to 4095) to

investigate the hardness of the objects as described in Section 1.3. The corresponding fingers' position readings from each force level and the proximity from the IR sensor are stored in the pressure-squeeze_n.csv files.

The vibration data is recorded in mV with a sampling rate of 400kHz. Immediately after every object's exploratory procedure, the background noise is recorded and stored under the folder called "background". The background sample is recorded while the NAO robot is performing exploratory procedures, and the RH8D is in the same position as if it were to hold an object. This is done to compensate in case the noise characteristic from the robots' actuators, cooling fans, and other moving parts changes during data collection. The vibration data was collected using the five channels/microphones in the positions described in Section 2.4.1 and stored in CSV files. There are individual files for each exploratory procedure ("feel" and "pressure-poke") and channels for three different object repositionings resulting in 30 files for vibration data.

The folder structure of the dataset is as follows:

```
Data
---|object
------|background
------|observation
---------|haptic_feel_noise
------------|Tactile_feel_n_ch_m.csv
---------|haptic_feel_clean
------------|Tactile_feel_n_ch_m.csv
---------|haptic_pressure-poke_noise
------------|Tactile_pressure-poke_n_ch_m.csv
---------|haptic_pressure-poke_clean
------------|Tactile_pressure-poke_n_ch_m.csv
---------|kinesthetics
------------|enclosure_n.csv
------------|pressure-squeeze_n.csv
------------|unsupported_holding_n.csv
------------|extra_n.csv
------|visual
--------|position-n.jpg
```

References

[1] Ender-3 Pro 3D Printer. https://www.creality.com/goods-detail/ender-3-pro-3d-printer.

[2] NAO the Humanoid and Programmable Robot – SoftBank Robotics. https://www.softbankrobotics.com/emea/en/nao.

[3] RH8D Adult size Dexterous Robot Hand. https://www.seedrobotics.com/rh8d-adult-robot-hand.

[4] Daniel Daugaard Buhl and Lasse Emil Ranulph Bonner. *Biologically Inspired Multimodal Fusion Learning for Robotic Object Recognition.* MSc, Department of Electrical and Computer Engineering, Aarhus University, Aarhus, Denmark, June 2021.

[5] J.E. Carlson, J. van Deventer, A. Scolan, and C. Carlander. Frequency and Temperature Dependence of Acoustic Properties of Polymers Used in Pulse-Echo Systems. In *IEEE Symposium on Ultrasonics*, volume 1, pages 885–888, Honolulu, HI, USA, October 2003. IEEE. 10.1109/ULTSYM.2003.1293541.

[6] Kristian Kristensen. *Towards Better Multimodal Object Recognition.* MSc, Department of Electrical and Computer Engineering, Aarhus University, Aarhus, Denmark, June 2021.

[7] Susan J. Lederman and Roberta L. Klatzky. Hand Movements: A Window into Haptic Object Recognition. *Cognitive Psychology*, 19(3):342–368, July 1987. 10.1016/0010-0285(87)90008-9.

[8] Susan J. Lederman and Roberta L. Klatzky. Haptic Perception: A Tutorial. *Attention, Perception, & Psychophysics*, 71(7):1439–1459, October 2009. 10.3758/APP.71.7.1439.

[9] Sibel Toprak, Nicolás Navarro-Guerrero, and Stefan Wermter. Evaluating Integration Strategies for Visuo-Haptic Object Recognition. *Cognitive Computation*, 10(3):408–425, June 2018. 10.1007/s12559-017-9536-7.