YOUNG EXPRESSIONS

EXPLORATORY WORK

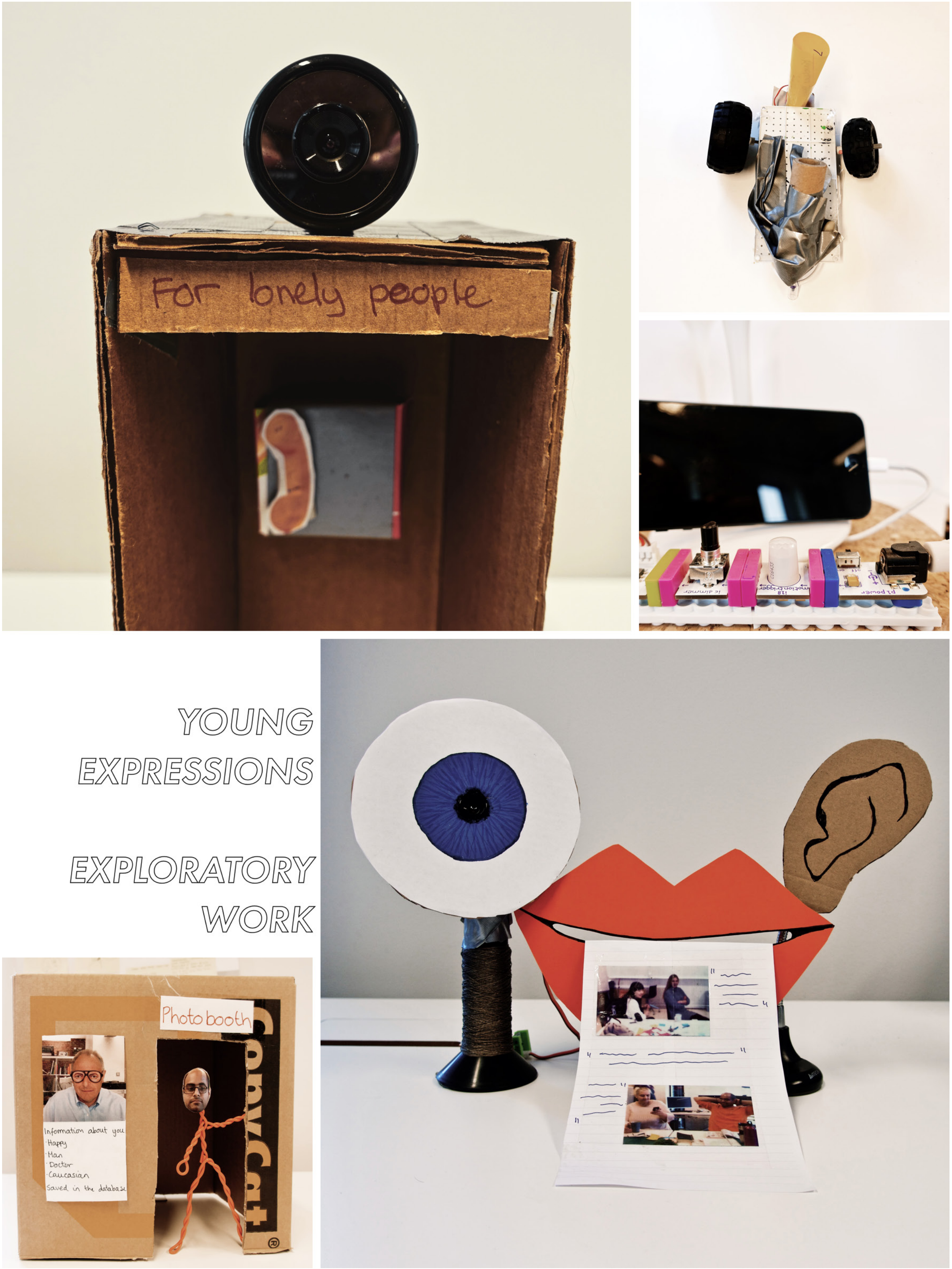

Figure: Low-fidelity prototypes and concept explorations from the Hearsay project.

Text Excerpt from:
Pandey, S. (2019). *Designing Smart Things* (Ph.D. Thesis,
University of Oslo).
Retrieved from http://hdl.handle.net/ 10852/68426

Process/Outcomes

To understand smart technology as a material in design, I attempted to build an understanding and sensibility of the material properties of machine learning through exploratory design rather than using it for a predetermined purpose or problem (Wiberg, 2014). My initial experiments were intended to explore lines of material centric inquiry, such as*, how do machines learn, infer, and interpret? what are the formative technological elements that constitute machine interpretations? how do the interpretations vary based on the nature of machine learning algorithms?* Within the context of smart consumer technologies, machine learning usually employs neural networks and deep learning algorithms which (currently) take the form of supervised learning. Within the context of smart consumer technologies, machine learning usually takes the form of 'supervised learning'. Supervised learning algorithms learn from 'labelled' data-sets that consist of pre-selected inputs and their true (or correct) output labels. The algorithm attempts to infer the relationship between inputs and outputs by predicting the output for a given input and identifying the error between the predicted and the true output label. The error is minimised iteratively based on pre-configured parameters, that strengthen specific correlations between input data and predicted labels and allow the algorithm to make progressively better predictions. For example, an algorithm can be trained on a large data-set, consisting of labelled images of different objects. The algorithm would attempt to correlate *particular object characteristics* in the images, to their true labels. Over multiple iterations, it identifies correlations that minimise the error. Once 'trained', the algorithm could potentially identify different objects in images outside the dataset as well.

Supervised learning is typically used for classification (into a predefined finite set of categories) or regression (inferring the relationship between a set of input and output values) related tasks. For example, automatically identifying and differentiating between people and/or objects in images is an example of classification while predicting the thermostat setting based on the time of day, season and weather is an example of regression. While these concepts helped outline a general understanding of machine learning and aided in the formulation of a concrete scope for my explorations, they were still quite abstract from a materials perspective. From the standpoint of machine learning algorithms, I found two different directions that could be explored — 1) neural network algorithms with predefined structures and large labelled data-sets, or 2) interactive machine learning programs that use humans as a part of the learning process to observe and modify inputs incrementally to improve the learning outcomes. Both of these techniques work for supervised learning tasks like image recognition but vary in their configuration process, flexibility, and accuracy, and therefore, have their own strengths and limitations. Implementing neural network algorithms allowed me to get a real-world understanding of machine learning along with getting more accurate results 'out of the box,' while interactive machine learning programs helped quickly explore different kinds of input-output relationships and work with very little dependency on large and external data-sets. I decided to try both approaches and conducted three exploratory experiments looking into different forms of machine interpretation that used audio and visual data streams as inputs. I started by looking at image classification and conversation modelling – two areas that have recently garnered a lot of interest within the machine learning community. These algorithms have also been integrated within some of the most popular forms of smart consumer technology, such as smart cameras and voice-based speakers and conversational interfaces, like chatbots. In my experiments, I used open-source software rather than web-based APIs to better understand the internal mechanics and complexity of the algorithms involved. I introduce the experiments briefly below –

1. *See/ML* (object recognition and scene interpretation): used a neural network to determine the objects in each frame of a video stream and then interpret the image textually based on the objects detected. I used an implementation of the 'neuraltalk' neural network algorithm (Karpathy & Fei-Fei, 2017; Karpathy, 2016) to generate the textual interpretations of the images.

2. *Hear/ML* (voice-based conversation generation/interpretation): used live audio streams as an input to create voice-based responses. I used an

implementation of the seq2seq neural network-based conversation modelling algorithm (Sutskever et al., 2014) in this case along with the python speech recognition library (Uberi, 2017).

3. *Emot/ML* (emotion/gesture interpretation): used movement cues and distance to judge the intent and emotional state of the person in front of it. I used a popular program for interactive machine learning called Wekinator (Fiebrink, Trueman, & Cook, 2011) and trained it on movie clips and camera feeds to experiment/explore new and wholly subjective forms of machine interpretation. Each study was intended to explore the allowable variations in the nature of the data stream and its implications on the nature of interpretations. The studies also involved varying the datasets (like in the previous cases) and the creation of new, smaller and non-comprehensive data sets.

The interpreted outcome from all three explorations was highly *generative* in nature. For instance, new sentences were generated during image interpretation using 'neuraltalk' that were not present in the training data-set. It highlights the fact that the algorithm does not merely remember the image descriptions from the training data but *infers* the relationship between the description and the objects in an image. Similar results were seen while using 'seq2seq' and 'DenseCap' as well where the program responded with entirely new responses during conversations and new image captions respectively. The generativity of the outcomes also points to the inherent *adaptability* of machine learning algorithms. Rather than remembering specific outputs correlating to the inputs in the training data-set, the algorithms infers relationships at a granular level, like the relationship between the words in a sentence (to be able to construct new and meaningful sentences). This allows them to reasonably adapt to a wide variety of new inputs that share *some* patterns of similarity with the training data set. This is different from rule-based systems that work with hard coded rules and consequently fail if the presented data does not match any of the rules. For example, 'seq2seq' was able to create responses to completely new and arbitrary input dialogs while 'neuraltalk' was able to generate an interpretation for new images or video streams. Generativity and adaptability can broadly be seen as aspects of the *extensible* nature of machine learning. Extensibility meshes well with the continuous nature of smart consumer technology discussed in the previous section, where the input data-stream, being situated in everyday life, may be quite unpredictable and varied. I think that it is the generative, adaptive and extensible nature of machine learning, that makes it hard to perceive and predict its *seams* (Chalmers & Maccoll, 2003) as a material. Machine learning can, therefore, be seamlessly and invisibly integrated into larger systems. However, even though it's hard to perceive, the seams do exist and often get highlighted in erroneous interpretations and examples of breakdowns from everyday life interactions [for instance, see (Hill, 2013; Zhang, 2015)]. Hear/ML and See/ML directly informed the next set of outcomes, that took the form of more tangible, albeit speculative, artefacts. Emot/ML, on the other hand, was important for helping understand the generative and adaptable nature of the interpretations generated by machine learning algorithms, in addition to helping qualitatively appreciate how the nature of data can affect the interpretations by biasing it in particular ways. However, I chose not to develop it further in design projects due to time constraints and since it did not seem stable enough to be used in everyday settings[17].

Due to my earlier work with the Hearsay concept and the notion of (visible and invisible) 'machine participation' in everyday lives, I focused on popular kinds of smart consumer technology for more concrete explorations. In addition, I felt designing personal artefacts could help explore alternatives to the existing ways in which (smart) consumer artefacts mediate everyday experience and practices through their material presence and materiality. Hearsay and the examples of breakdowns in everyday interactions with smart speakers (Pandey & Culén, 2017), became a point of departure for me for further explorations with conversation modelling (Hear/ML). While thinking of alternate and more personal forms of interaction and presence, I also started to think of the dominant forms of interaction that can be seen in smart consumer technology, such as the absence or minimal presence of manual controls and their dependence on smartphone applications for control and configuration. As I attempted to conceptually frame and explore my work further, the counterfunctional framing[18] (Pierce & Paulos, 2014a) helped situate it in a larger design space and evolve it by emphasising alternate values and forms of presence.

Hearsay's physical form draws inspiration from the playful yet striking aesthetic of Italian radical design (Malpass, 2017), specifically that of the Memphis-Milano design group (http://memphis-milano.org) and Studio Alchimia (http://www. alchimiamilano.it/). The removable cover (lampshade) is translucent and shows a faint outline of an evocative physical form inside. The evocative form contrasts with the minimal cover, giving Hearsay a layered aesthetic. Functionally, Hearsay is a lamp, which can be switched on and off using voice commands. If the removable cover of the lamp is kept on (covered state), the audible responses are muted (but are still generated and saved) and the interactions are limited to controlling the lamp. Removing the cover, un-mutes the artefact and reveals the evocative form (uncovered state). The form is used to highlight the artefact's material composition, like the speaker, microphone, network connection, and a transcript of all the conversations and responses (captured both while muted and unmuted). In the uncovered state, the light from the lamp is also dimmed to create a soft and intimate environment for conversations. Hearsay connects to the internet using a pre-configured wireless router that needs to be attached to the user's modem via an Ethernet cable. Hearsay

[17] Wekinator (Fiebrink, Trueman, & Cook, 2011) is more applicable for settings where variations in the form of interactions are relatively controlled like in the case of generative music experiments, such as Spring Spyre (EAVIgoldsmiths, 2014). While I could not explore it further in my own work, it was used by two groups of students in a course I was involved in, exploring alternate expressions of smart consumer technology. The functional characteristics of their concept required the algorithm to differentiate between three sets of facial expressions and since their prototype was just required to be a proof of concept, Emot/ML seemed like a quick starting point for exploration and implementation.

[18] Pierce and Paulos describe counterfunctional things as "a thing that figuratively counters some of its own 'essential functionality'" (Pierce & Paulos, 2014a). They suggest removing, inhibiting, and/or inverting essential functionality to define alternatives can support critical reflection on what exists while also presenting new and sometimes overlooked possibilities and opportunities for design.

automatically connects to the internet once the router is attached. Once connected, Hearsay is always listening and responds as soon as it detects audible and discernible voices. This is in opposition to most voice-based interfaces, that get activated using a hotword [a particular keyword like 'Alexa' (https://developer.amazon.com/alexa), 'OK Google' (http://bit.ly/okgoog), etc.].[19]

Besides helping understand machine learning as a material, Hear/ML helped me speculate about alternate forms of interactions and presence of smart speakers (and smart technology in general) in everyday life. For instance, due to its adaptability, the seq2seq algorithm could generate responses for a wide variety of conversation snippets. While in real world applications, it is quite unstable and often starts falling back to complete gibberish, but conceptually, it let me define a 'continuous listening and response' based interaction for Hearsay. In contrast, a rule-based system, would only have responded to pre-configured commands and presented an error in other cases. Moreover, it would have reduced the serendipitous nature of the interaction by generating predictable responses. Hearsay's surprising and casual interaction was largely a factor of it being trained on a 'non-utilitarian' movie subtitle dataset (Danescu-Niculescu-Mizil & Lee, 2011), rather than a dataset of more 'routine' conversations.

[19] This description of Hearsay's functional characteristics is adapted from the paper, *Framing Smart Consumer Technology: Mediation, Materiality, and Material for Design* (Pandey, 2018b)