HEARSAY: SPECULATIVE EXPLORATION OF INTELLIGENT VOICE BASED INTERFACES

ARTICULATION
SUMIT PANDEY
UNIVERSITY OF OSLO
SUMITP@IFI.UIO.NO

NEGOTIATIONS
ALMA LEORA CULÉN
UNIVERSITY OF OSLO
ALMIRA@IFI.UIO.NO

ABSTRACT

In this paper, we present a reflective visual account of the process and outcome from a speculative research through design project – Hearsay. Through this account we unpack and present the conceptual, technical and material explorations that guided our design process. Further, using this mode of reflective visual articulation, we contribute to interaction design research by highlighting potential possibilities and problematics for design within the emergent space of intelligent voice based interfaces.

INTRODUCTION

With the recent push towards intelligent voice based interfaces in everyday objects like phones (Apple 2017), speakers (Amazon 2017a; Google 2017) and even refrigerators (Cunningham 2017), the practices of interaction design need to adapt to the new space of 'intelligent' networked devices. Being new and relatively unexplored, designs from the mass market are largely limited to interface explorations or control and query based feature designs. However, to extend our understanding of the possibilities and problematics latent within this space, we argue that there is a need for more exploratory and speculative engagement with it removed from the constraints of the mass market.

In this paper, we present the process and outcome from such an exploratory and speculative (Auger 2013) research through design (Zimmerman and Forlizzi 2014) project – 'Hearsay'. Hearsay is an 'always-on' and 'always conversing' voice activated lamp that generates quirky and whimsical rather than efficient and functional responses while incorporating transparency rather than obscurity in its form and design (Figure 2).

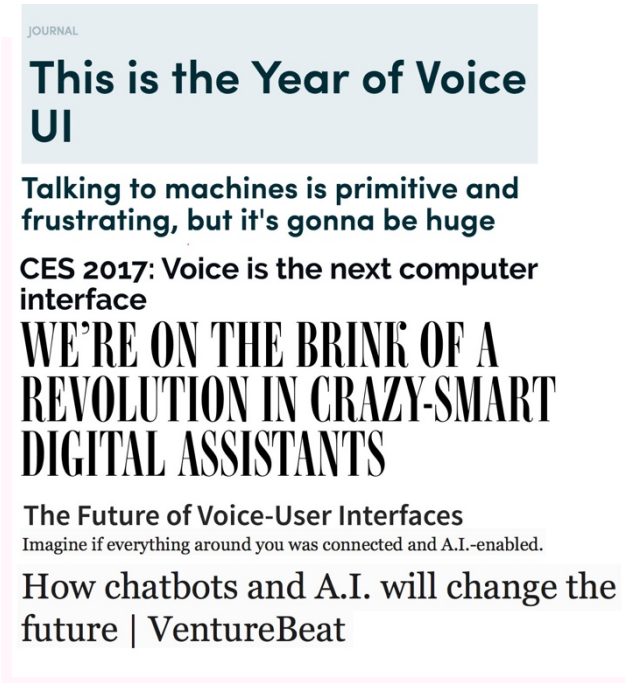


Figure 1: Intelligent Voice based Interfaces in mass media.

Our intent behind unpacking the design process, speculative explorations and outcome is twofold. By sharing a rich visual account of our exploratory process and its outcome, we wish to highlight the relatively unexplored possibilities of interaction within this space along with specific issues and problematics present within it. Further, we intend to extend the understanding of the complex and often obscured (Knutsen 2014) 'intelligent' technologies from a design perspective by discussing and unpacking the algorithms and systems that power Hearsay.

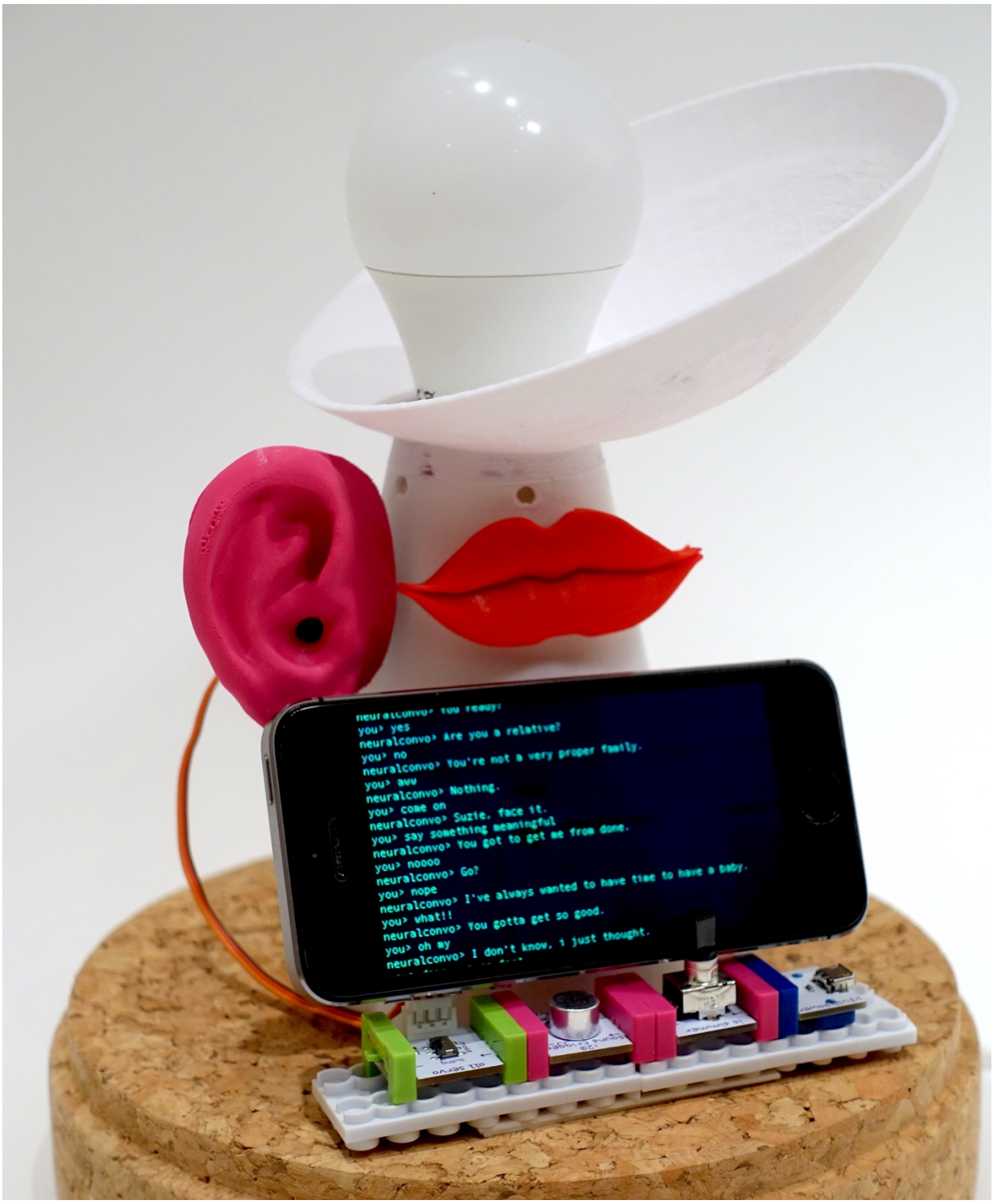


Figure 2: Hearsay –an ‘always-on’ and ‘always conversing’ voice activated lamp that generates quirky and whimsical rather than efficient and functional responses. The conversations are immediately transcribed and are always available for the owner to see.

THE DESIGN PROCESS

CONCEPTUAL EXPLORATION

The design process for the project took the form of parallel conceptual and technological explorations. Since largely, the current discussions on natural voice-based and conversational interfaces in the mass media represent them as the next frontier of interface/interaction/experience design (Kojouharov 2016; Newman 2016; Tuttle 2016) (see Figure 1) it made us interested in the *fringes* of mass media reports.

These reports discussed the commonplace fears and cultural understanding of these interfaces along with the implications of glitches and unfulfilled promises (see Figure 3). This also allowed us to develop an articulation of the environment and context in which these devices are being experienced. Both Auger and Dunne and Raby emphasise the importance of grounding the speculative artefact in a "logical reality" and highlight that it may "appear odd" at the outset but challenges preconceptions and allows for alternative perspectives and understandings of technology to emerge (Auger 2013; Dunne and Raby 2013).

We collected these fringe reports in the format of a continuously evolving physical and online collage (Figure 3) and used it to uncover problematics in this design space in addition to using them as seeds for probable design concepts and 'what ifs?' situated in 'possible worlds' (Dunne and Raby 2013: 90–93).

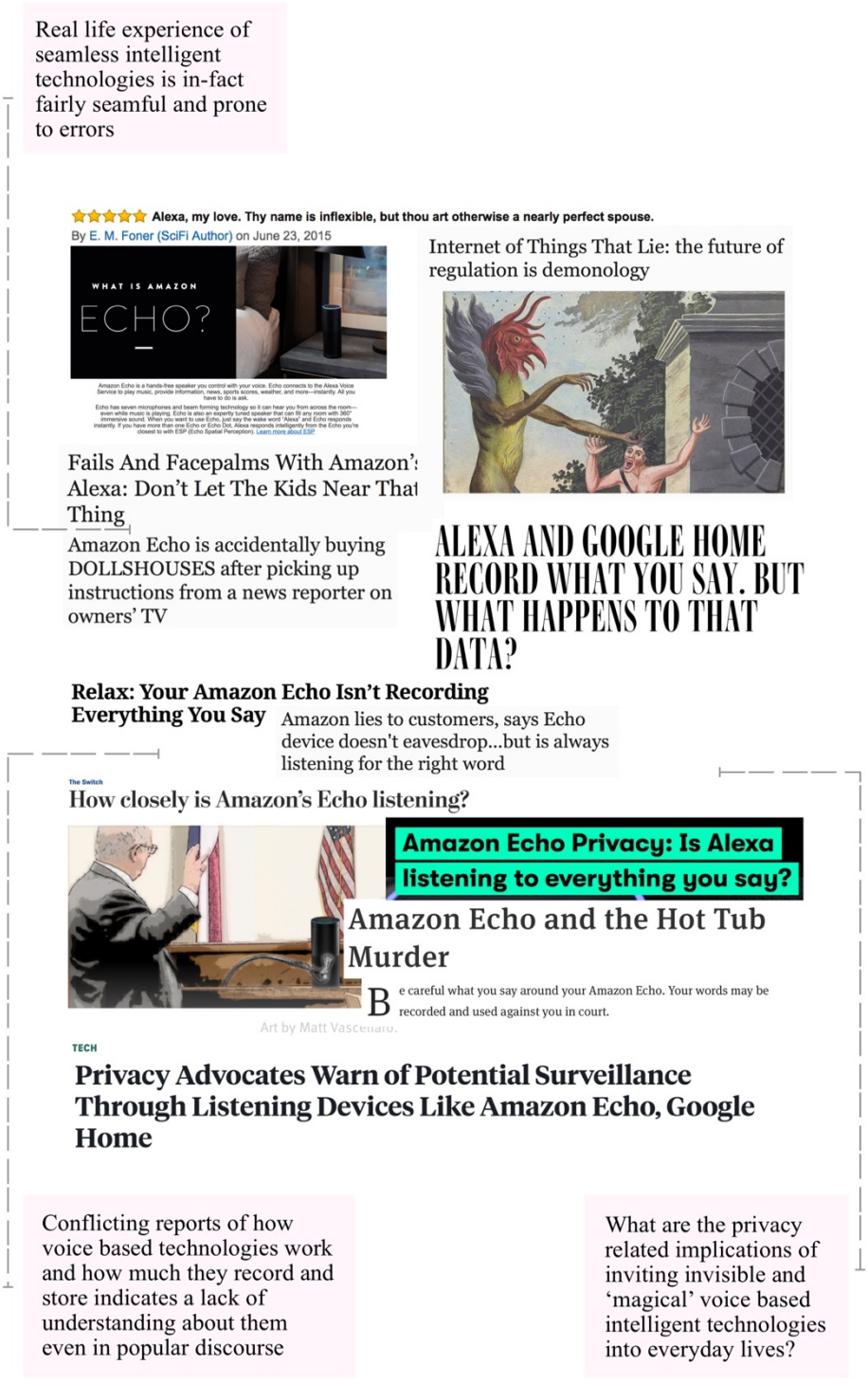
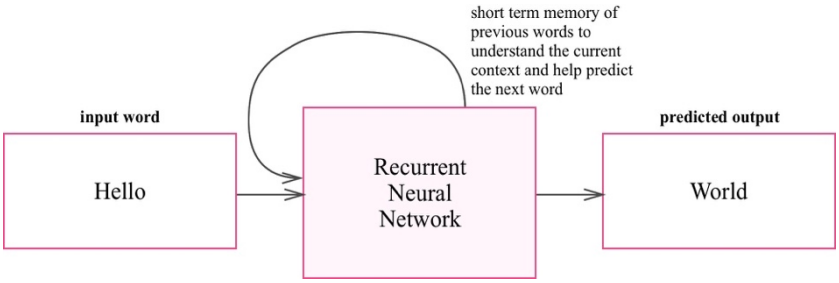


Figure 3: Mass media reports on the fears and misunderstandings of Intelligent Voice based Interfaces.

TECHNOLOGY EXPLORATION

From a technology standpoint, we started to explore different functional aspects of voice based virtual assistive devices like ‘always-on hot-word detection’, ‘voice processing and response’ and realised that the technology used to generate ‘conversation responses’ was based on a neural network algorithm called Recurrent Neural Networks (RNN) (Figure 4).

Figure 4: A RNN predicting the most likely word that could follow the input ‘Hello’ (a similar system is present in most current generation smartphones).



While an in-depth discussion of the technical aspects of neural networks and RNNs is out of scope for this paper, we would like to introduce some fundamental concepts (with the risk of over-simplification) for the sake of a clearer discussion and understanding of our choices and un-black-boxing Hearsay’s functional aspects. Neural networks refer to a type of machine learning algorithm that *infer and build* a predictive or decision making logic based on a dataset of input output relationships, rather than following a set of predefined/hardcoded rules (Gershenson 2003). Recurrent neural networks are a special class of neural networks that infers relational patterns between words and characters rather than learning specific responses to fixed queries. Consequently, it can generalise its learning to new questions and cues that are not present in the training dataset as well (Vinyals and Le 2015). This makes them quite effective for conversational modelling and we decided to use the seq2seq (Vinyals and Le 2015)/neuralconvo algorithm (Cournoyer 2017) locally for generating responses to voice cues (Figure 5, top - below).

While the functional quality of responses generated by a locally implemented program is not the same as that from a cloud API (like Amazon (2017b) Alexa), it gave us the freedom of being able to experiment and explore the nature of responses that could be generated by different data sets. The algorithm was finally trained on a dataset of movie subtitles both because of ease of access and the casual and informal nature of responses generated. However, for reasons of accuracy, a cloud based API was used for the speech to text function which was then processed locally by the RNN mentioned earlier.

Human: *what is moral?*
Machine: *what empowered humanity, what intellectual the essence is.*
Human: *what is immoral?*
Machine: *the fact that you have a child.*
Human: *what is morality?*
Machine: *what is altruism?*
Human: *what is the definition of altruism?*
Machine: *if you don’t believe in god, then you don’t know.*
Human: *ok ... so what is the definition of morality?*
Machine: *well, the truth is, you’re not a believer in god almighty.*

```
Human: hello
Machine: you ready?
Human: yes
Machine: are you a relative?
Human: no
Machine: you're not a very proper family
Human: aww, come on
Machine: suzie, face it.
Human: say something meaningful
Machine: you got to get me from done.
Human: nope
Machine: I've always wanted time to have a
baby.
Human: What!
Machine: You gotta get so good.
```

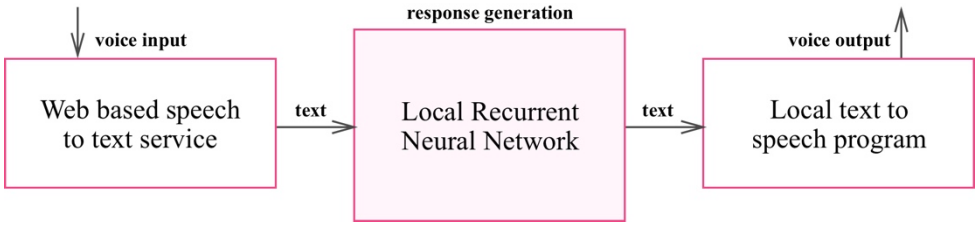
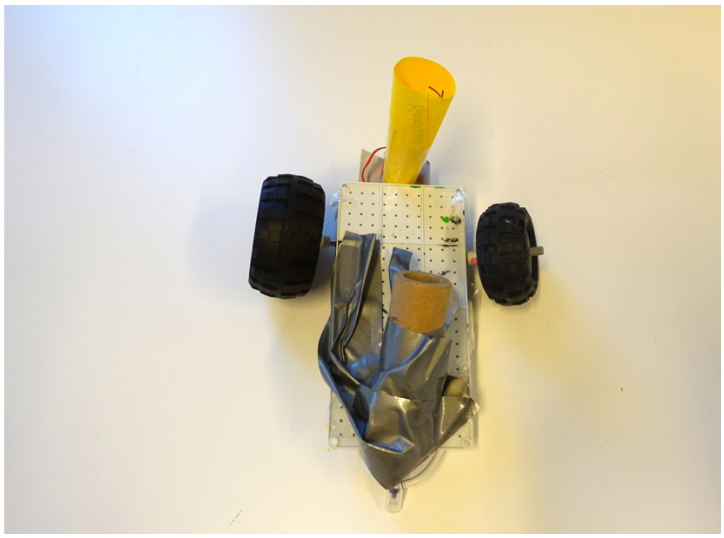


Figure 5: Human-machine conversations. From (top - above) ‘a neural conversation model’ (Vinyals and Le 2015) and (top - below) a locally deployed neural network trained on movie subtitles. (Left) the final system diagram for Hearsay highlighting the technologies used

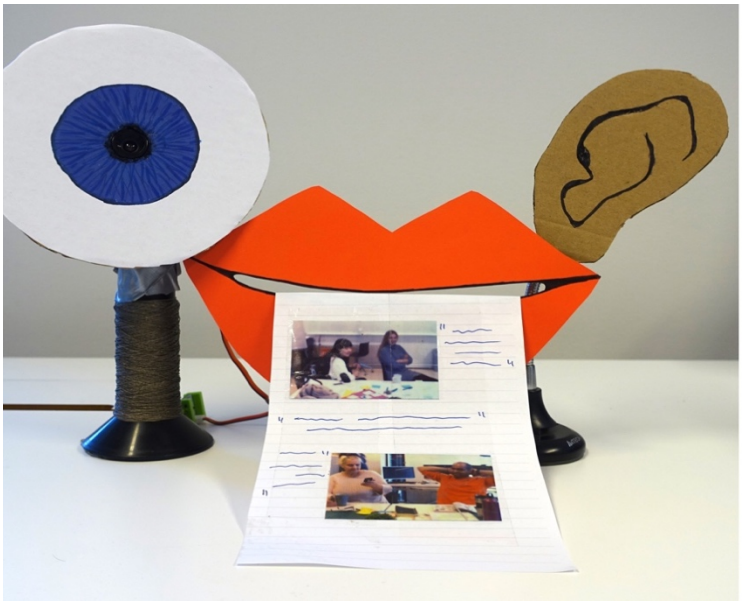
EARLY PROTOTYPES

Based on the speculations and problematics outlined during the conceptual explorations and findings from the technological explorations, concept sketches followed by quick low-fidelity prototypes were generated to serve as material props that were used for greater engagement during discussion and critique.

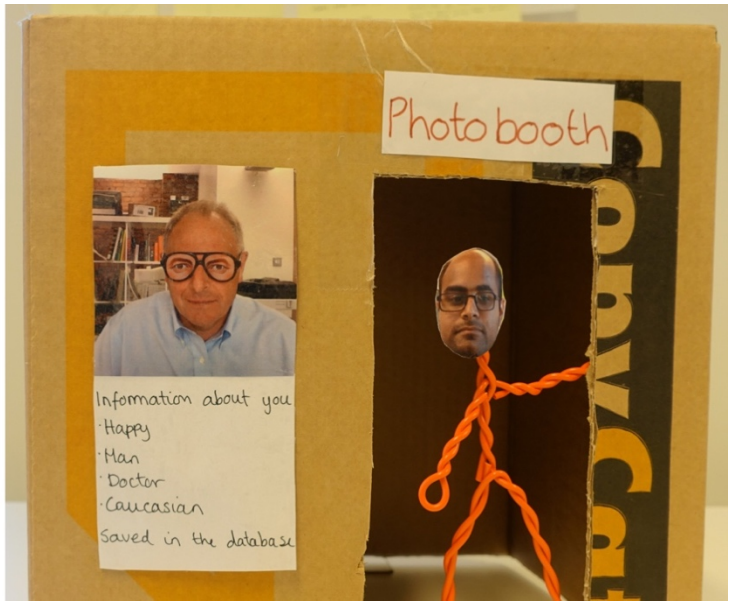
The lonely booth: automatically detects lonely people using emotion detection and tries to talk to them.



Voice based drawing machine: drawing machine that interprets voice commands and emotion and draws accordingly.



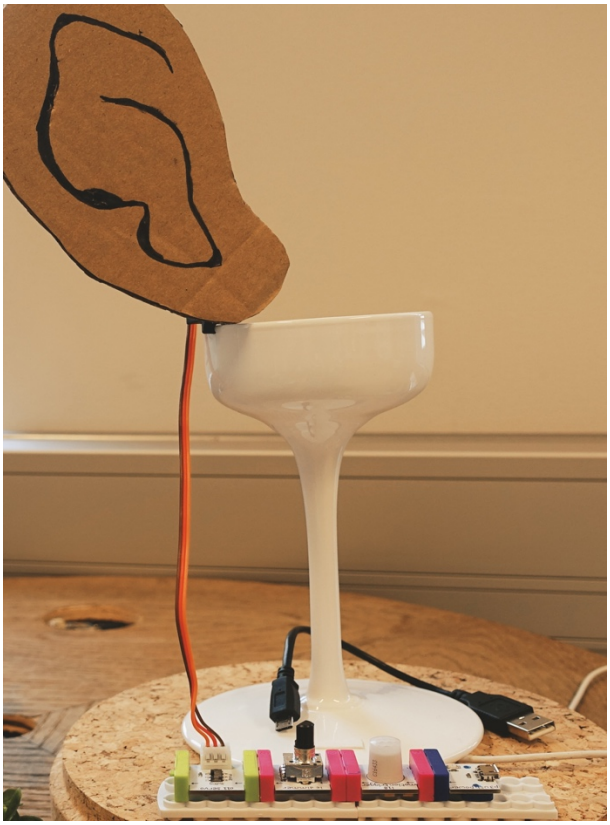
The God eye/mouth/ear: always on recording, interpreting and sharing machine for public spaces, (a precursor to Hearsay).

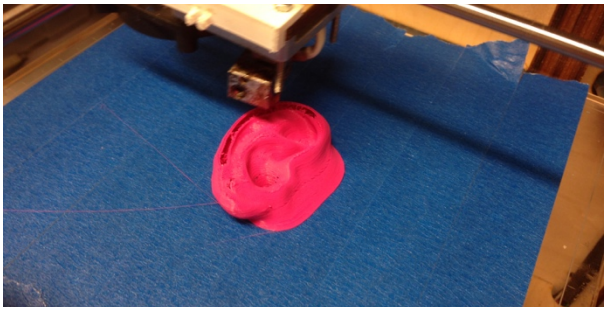


The photo booth: Autonomous photo-booth that interprets an image and prints the closest possible match from the internet, based on a textual interpretation.

FORM & INTERACTION

Creating interactive prototypes and formal sketches helped critique design explorations while also showing new or less explored design dimensions during the process. Using this combination of exploration and reflection was central to our process, highlighting a baseline for outlining a ‘likely’ and appropriate (but not stereotypical) form for Hearsay along with its behaviour and function.





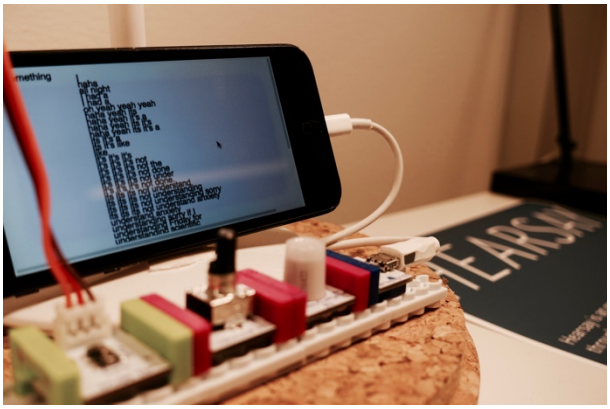


Figure 6: (Left) Moodboards and concept and form sketches, (Right) Initial form and technology explorations

HEARSAY

Hearsay is an intelligent networked lamp that uses conversation as a means of interacting with the world. Considering the ecology of a domestic household, it was designed to take the external form of a mundane domestic object (a lamp), complemented with always-on networked intelligence. However, the aesthetic of the final artefact was deliberately designed to be layered – seemingly mundane and domestic on the outside and personal, provocative and communicative on the inside (Figure 2, 7). While the intended effect was to create viewer interest and invite exploration and discovery, it was also meant as an expression of the opposition between the modern, neutral and impersonal aesthetic of other voice enabled technological devices and the alternative reality presented by Hearsay.

From a utilitarian standpoint, Hearsay's lamp function can be switched on and off using voice commands. However, rather than responding to utilitarian queries with functional responses, Hearsay tries to have a real conversation with its owner (see Figure 9). This may not always be possible since while the external cover of the lamp is kept on, the audible responses are muted (but the conversation is still active and the responses are simply transcribed instead) and the interactions are limited to controlling the lamp.

Removing the external cover uncovers evocative forms that are used to communicate the artefact's internal and external functions. It also enables audible responses and exposes a transcript of the ongoing conversation that the device has been carrying on with the owner (albeit silently).

Figure 7: Hearsay in a studio exhibit setting

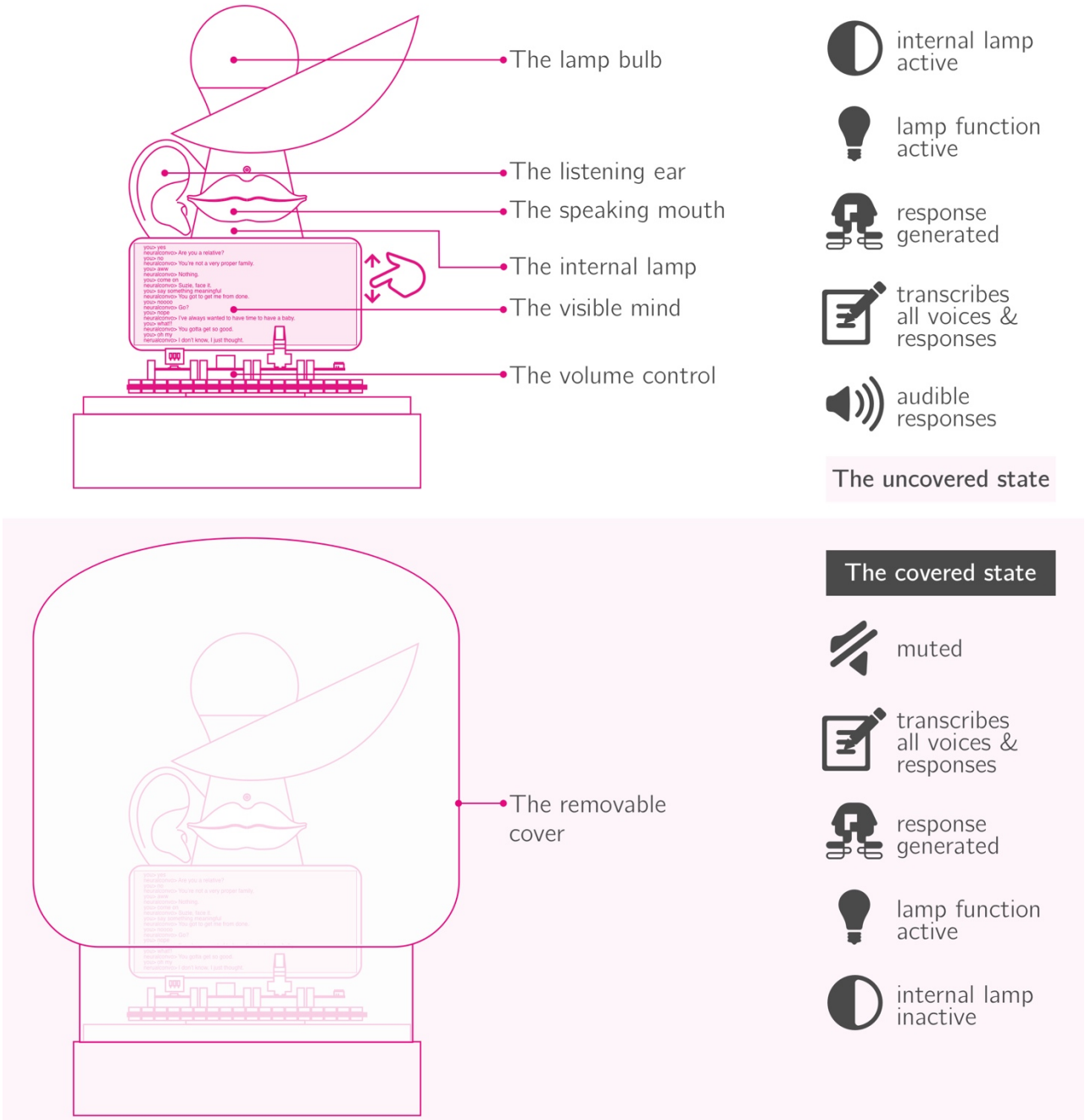


Figure 8: Functional Elements

HEARSAY: FUNCTIONAL ELEMENTS

The external cover of the lamp can be removed to reveal the sparse internals of the machine - an internal lamp that offers a much softer glow, the listening ear, the speaking mouth and the visible mind (Figure 8).

1. *The visible mind:* is a screen that highlights a transcript of all the conversations that have happened.
2. *The listening ear:* encapsulates a microphone and moves whenever the machine is recording.
3. *The speaking mouth:* encapsulates a speaker using which the machine responds to its owner. Activated only in the uncovered state.
4. *The internal lamp:* is a secondary light that switches on in the uncovered state and disables the main lamp. It serves as a counterpoint to the harsher main light by offering a soft and intimate light more suited to having a conversation.

Place hearsay next to your bed and have wonderful conversations late into the night!

Jane: silence and boredom, not good bed-fellows.
Hearsay: well, you can go on.
Jane: even wine isn't helping tonight
Hearsay: Lonnie, I think it is!

Hearsay is always on and always in conversation, even when covered. Uncover it and enjoy all the quips and remarks you missed.

Jane: what should I cook? I need something nice!
Hearsay [muted]: You're one hell of a detective, Mrs. Christian :>

Don't need any conversation, just the light? Pop the lid back on and Hearsay will go silent (but still listening)

John: Turn on the lights.
Hearsay [muted]: sure, but you should call her.

Enjoy having a conversation as you open hearsay up for changing the bulb or just to see what its been thinking of.

John: Ah, looks like I need to head out to the store today.
Hearsay: I knew you were a cop but you never believe me.

ALWAYS IN CONVERSATION.

Hearsay is always engaged in a conversation with you and understands that single word confirmations may not be the best way to engage.

HEARSAY

Figure 9: Promotional scenarios of use. The text advertised consists of real responses from Hearsay.

Figure 10: Advertisement poster for Hearsay

CONCLUSION

Voice based intelligent networked technologies present new opportunities and challenges for interaction design research and practice. A useful way of understanding these implications (beyond utopian and dystopian visions) is through reflective engagement with the materials of this new space. By describing the technical, conceptual and formal aspects of our process and outcome, we have tried to highlight the possibilities and problematics linked to material and conceptual exploration latent in this rapidly evolving design space and how this form of material engagement can lead to a diverse range of inquiries. By incorporating elements of transparency and discovery both as a formal and an interactive quality, we have tried to explore Hearsay as a layered and personal artefact. While it does highlight the extent of machine intrusiveness in everyday life, it also intends to make gaps in machine understanding and interpretation more transparent and creates a reflective space around humanized and always on forms of artificial intelligence.

We see Hearsay, and our visual account outlining its process and intent, as a *particular and concrete* exploration of many possible speculative and material lines of inquiry in this design space. Rather than as a finished usable/useful product we would like it to be read and experienced as an unfinished proposal (Sengers and Gaver 2006) that raises questions, concerns and propels designers' and researchers' imagination with new forms of material understanding and curiosities. Therefore, we suggest that our reflections and speculations be read as snapshots of our understanding and intent and a point of departure for further investigations, rather than as final, concrete outcomes and best practices.

ACKNOWLEDGEMENTS

This project was, in part, financed by the EU Creative Europe project 'The People's Smart Sculpture', under the grant number EC-EACEA 2014-2330. The authors would also like to thank Jorun Børsting and Michelle Cheung for their help with the initial concept and prototype development.

REFERENCES

Amazon (2017a) Amazon Echo [WWW Document]. URL https://www.amazon.com/Amazon-Echo-Bluetooth-Speaker-with-WiFi-Alexa/dp/B00X4WHP5E (accessed 1.16.17).

Amazon (2017b) Alexa [WWW Document]. URL https://developer.amazon.com/alexa (accessed 1.16.17).

Apple (2017) iOS 10 - Siri [WWW Document]. Apple. URL http://www.apple.com/ios/siri/ (accessed 1.16.17).

Auger, J. (2013) Speculative design: crafting the speculation. Digit. Creat. 24, 11–35. doi:10.1080/14626268.2013.767276

Cournoyer, M.-A. (2017) macournoyer/neuralconvo.

Cunningham, A., 2017. LG threatens to put Wi-Fi in every appliance it introduces in 2017 [WWW Document]. Ars Tech. URL http://arstechnica.com/gadgets/2017/01/lg-puts-amazons-alexa-in-a-fridge-and-wi-fi-in-everything-else/ (accessed 5.16.17).

Dunne, A., Raby, F. (2013) Speculative Everything: Design, Fiction, and Social Dreaming, 1st edition. ed. The MIT Press, Cambridge, Massachusetts ; London.

Gershenson, C. (2003) Artificial neural networks for beginners. ArXiv Prepr. Cs0308031.

Google (2017) Speech API - Speech Recognition | Google Cloud Platform [WWW Document]. Google Dev. URL https://cloud.google.com/speech/ (accessed 1.16.17).

Knutsen, J. (2014) Uprooting products of the networked city. Int. J. Des. 8.

Kojouharov, S. (2016) How chatbots and A.I. will change the future [WWW Document]. VentureBeat. URL http://venturebeat.com/2016/07/24/how-chatbots-and-a-i-will-change-the-future/ (accessed 10.20.16).

Newman, D. (2016) Chatbots And The Future Of Conversation-Based Interfaces [WWW Document]. Forbes. URL http://www.forbes.com/sites/danielnewman/2016/05/24/chatbots-and-the-future-of-conversation-based-interfaces/ (accessed 10.20.16).

Sengers, P., Gaver, B. (2006) Staying Open to Interpretation: Engaging Multiple Meanings in Design and Evaluation, in: Proceedings of the 6th Conference on Designing Interactive Systems, DIS '06. ACM, New York, NY, USA, pp. 99–108. doi:10.1145/1142405.1142422

Tuttle, T. (2016) The Future Of Voice-Activated AI Sounds Awesome [WWW Document]. TechCrunch. URL http://social.techcrunch.com/2015/03/06/the-future-of-voice-activated-ai-sounds-awesome/ (accessed 5.24.17).

Vinyals, O., Le, Q. (2015) A Neural Conversational Model. ArXiv150605869 Cs.

Zimmerman, J., Forlizzi, J. (2014) Research Through Design in HCI, in: Olson, J.S., Kellogg, W.A. (Eds.), Ways of Knowing in HCI. Springer New York, pp. 167–189. doi:10.1007/978-1-4939-0378-8_8