SynKu: Exploring the Production of Sensory Objects

Ankur Agrawal  
Department of Human Centered Design & Engineering  
University of Washington  
ankurag@uw.edu

Mark Stamnes  
Department of Human Centered Design & Engineering  
University of Washington  
mstamnes@uw.edu

Wenvi Hidayat  
Department of Human Centered Design & Engineering  
University of Washington  
wenvi@uw.edu

Meishen Yin  
Department of Human Centered Design & Engineering  
University of Washington  
meisheny@uw.edu

Aravind Ravi  
Department of Human Centered Design & Engineering  
University of Washington  
aravi@uw.edu

Daniela K. Rosner  
Department of Human Centered Design & Engineering  
University of Washington  
dkrosner@uw.edu

Abstract  
This paper describes SynKu, a mobile application that merges audio and image processing with human interpretation to speculate on the production of sensory objects. Sensory objects are representations of sensory phenomena interpreted and used by both people and software. We discuss how the SynKu application enables us to explore how algorithms help to change and legitimate the reconstruction of sensory phenomena across self-tracking platforms.

Author Keywords  
Cross-sensory perception; science and technology studies; sensory objects; design inquiry.

ACM Classification Keywords  
K.4.0 Computers in Society: general.

Introduction  
Within HCI, a recent turn toward self-tracking has implicated smartphones in the tracing, codification, and analysis of myriad human activities, from exercise to sleep [2]. With this new attention to sensing, researchers rely on algorithmic reconstructions of sensory experiences that otherwise remain tacit, such as sights and sounds. As people move toward sharing this sensory data with others, it becomes increasingly
crucial to acknowledge and question the limitations of this reconstruction.

Debates on algorithmic accountability have recently emerged within the budding field of platform studies [1]. To examine the broad implications of this work, American media studies scholar Tarlten Gillespie [8] recently outlined six key aspects: patterns of inclusion, cycles of anticipation, evaluation of relevance, promise of algorithmic objectivity, entanglement with practice, production of calculated publics. Together, these dimensions describe central areas of investigation around the promise and purpose of algorithms within human knowledge practices. From recommendation systems to search engines, the procedures algorithms encode shape what forms of information become relevant. British sociologist Scott Lash [10:71] calls such activity generative - not only "open[ing] up opportunity for invention" but also constructing "pathways through which capitalist power works." Yet, we still know little about the form algorithms take as they shift and integrate with our sensory perceptions. Our project builds on this attention to the "platform" [1] through methods of design and intervention.

In the paper that follows we discuss our recent work to explore emergent forms of sensory perception with an Android application we developed called SynKu. SynKu processes images and sounds to create prompts for cross-sensory media recording (e.g., a “black and white” audio clip, a “high pitch” image). To do this, SynKu first processes the recorded images and audio to produce categorizations along with various spectra (saturation, lightness, and hue for images; volume and pitch for sound). Once processed, SynKu prompts the user to collect a different type of media according to

Figure 1: Screenshots of SynKu’s “Capture” mode. SynKu processes images and sounds to create prompts for cross-sensory media recording in the following sequence: a) SynKu prompts the user to capture a photo, processes the image and categorizes it as "bright," b) SynKu prompts the user to capture a sound that is "bright," processes the recorded sound and categorizes it as "high pitch," d) SynKu prompts the user to capture an image that is "bright," processes the image and categorizes it as "black and white," e) SynKu prompts the user to capture an image that is "black and white," processes the image and categorizes it as "soft," and f) SynKu stitches together the captured media to create video narrative or "Ku.”
that new category (e.g., "capture a sound that is bright" or "snap a pic that’s slow"). SynKu interweaves the algorithmic processing of five such media recording activities using ten media attributes (e.g. light / dark, loud / soft) with users’ cross-sensory interpretation of those attributes. People can then share the resulting media assemblage as a short video narrative (or "Ku"). By creating and sharing Kus, they produce new sensory objects interpretable by people and software. Here we discuss how this research enables us to examine the digital expression of sensory perception.

**Related Work**

A growing number of applications support the integration of audio and still image capture. Image audiography applications like Photoblab (https://www.photoblab.com/) and Speaking Photo (http://speakingphoto.com/) add sound to single images, enabling people to share additional aspects of the photographic situation. Speaking Photo further facilitates storytelling through image and sound sequencing and then allows users to share the resulting media in an online gallery. Social media based Vine (https://vine.co/) allows users to create short video loops with audio annotations while Julian Glander’s mobile application known as “Roy G Biv” [7] creates a synthesizer by reading hue, saturation, and brightness information from user supplied images. We extend these applications in three ways: firstly, we emphasize synesthetic mappings over isolated sensory experience; secondly, we use the resulting media recording as a provocation for social inquiry; and finally, we designed the application to examine the production of *sensory objects*. We conceptualize sensory objects as representations of sensory phenomena interpretable by people and software after Prentice [13], as we describe in more detail below.

Following on HCI’s tradition of using artifacts as conversational prompts [6], we focus on the mobile application’s role in social inquiry: enabling us to investigate how algorithms influence and legitimate the reconstruction of sensory phenomena across self-tracking platforms. Recent HCI projects have employed creative prompts as part of brainstorming activities. Faste, *et al.* [5] present chainstorming as a process by which people pass along design concepts to drive ideation. Desjardins, *et al.* [4] use techniques borrowed from the Surrealist game known as Exquisite Corpse to create a quick video sketching tool. For both, the process of introducing an element of ‘randomness’ in the connection of concepts supports the work of catalyzing novel designs. SynKu builds on the idea of creative prompting in the domain of sensory experience.

**SynKu**

As a genre of poetry, Japanese haikus consist of three units of five, seven, and five syllables in a verse form. According to neurologist Richard Cytowic [3], the celebrated Japanese haiku poet Matsuo Basho (1644-1694) readily used synesthesia to combine varied sense qualities. “It is an effective device for presenting an experience in its totality compared to the scientific method, which breaks the whole into its component parts,” Cytowic writes. “In part due to cultural forces, literary synesthesia in Japan presumes an attitude that accepts the ultimate interrelatedness of all things.”

Inspired by this connection, the SynKu application weaves together synesthetic prompts with a five-part
format reminiscent of the haiku. SynKu consists of mobile software that runs on an Android OS smartphone. Extending prior work on tools for synthethetic experience and sensory perception (e.g., [4][7]), we identify three core features of the SynKu system: gather, process, and share.

- **Gather:** SynKu produces five prompts for people to gather images and sounds. Each prompt changes according to the prior recording and encourages cross-sensory interpretation (e.g., a “high pitch” image or “black and white” audio clip). Once it collects the five media records, SynKu assembles them into a single video we call a "Ku" (Ku from the second syllable of Haiku).

- **Process:** SynKu processes the recorded images and sounds to generate each successive prompt.

- **Share:** SynKu shares the resulting Ku with others.

We designed the SynKu application to investigate the implications of sensory experiences legible to people and software. By first gathering images and sounds and subsequently sharing them with others, SynKu may inspire conversation about the digital workings of sensory perception. Using SynKu, we explore how algorithms extend and refigure sensory phenomena across self-tracking platforms. To understand how, consider the following scenario of use.

**Scenario of use**
Angela woke up to a sunny morning in Seattle. The light streamed through her windows and illuminated her room, inspiring her to open the SynKu app and snap a photo of the morning sunshine (Figure 1a). The app immediately processed the image, categorized it as bright, and prompted Angela to capture a ‘bright’ audio clip. Angela decided to walk to the park. Thrilled to play on the swings at the playground, the children at the park struck Angela as particularly bright, and she opened SynKu to record the sound (Figure 1b). Processing the sound and classifying the recording as high-pitched, SynKu then prompted Angela to capture a ‘high pitch’ image. After spending a long day out, Angela decided to stop for some ice cream before heading home. She chose chocolate and vanilla, squealing as she finished her cones. Finding this an appropriate ‘high pitch’ moment, Angela took a photo of the ice cream (Figure 1c). After processing the image, SynKu generated a new prompt: “record audio that’s ‘black and white’." Angela waited until her children were in bed before reopening SynKu. They like to get their dad to read a bedtime story to them, and ‘The Three Little Pigs’ is their favorite. Angela decided to record the storytelling session on SynKu (Figure 1d). The recording was then processed to provide a final

![Figure 2: Screenshots of SynKu's “Explore” mode. SynKu saves the media records as a single video narrative or “Ku” that people can share with others (left). By stitching together and sharing Ku’s, they produce new sensory objects interpretable by people and software (right).](image-url)
prompt: “snap a pic that’s ‘soft’.” After a long day, Angela ends her Ku with a photo of her partner falling asleep, feeling thankful to have her family by her side (Figure 1e). Angela built a sensory object in a Ku, ready to share and discuss with her family and friends.

How it works
The SynKu application is built on the Android platform (supporting Android OS 4.0-5.0). The application uses Parse API (https://www.parse.com) for backend operations such as login, storing and retrieving audio and image files. The application has image and audio processing modules for generating prompts.

The image-processing module generates prompts for recording audio using a histogram based heuristic - light or dark based on lightness, bright or dull based on saturation and colorful or black and white based on hue. To generate the prompt, the algorithm first downsamples the captured image and calculates the weighted average of three color channels, hue, saturation and value, using separate histograms. Next, the deviation is calculated for each of the channels from a threshold set for each of them, 5 for Hue and 127 for Saturation and Value. The deviation is then normalized to a 0-1 scale for comparison. The system then selects the channel with the maximum deviation and generates a corresponding prompt.

The audio processing module generates prompts for capturing images: soft or loud based on maximum amplitude and high-pitch or low-pitch based on frequency. The algorithm first processes the PCM (pulse code modulation) audio data to find the maximum amplitude in dB. Then the PCM data is processed for average frequency using a Fast Fourier Transform based YIN pitch tracking algorithm (a frequency estimator based on autocorrelation) from the Tarsos DSP library. The thresholds used for these classifications are 500Hz for pitch and 90 dB for loudness.

Our software saves the audio and image files along with the prompts on the Parse servers for subsequent engagement by other people using the SynKu app.

Future Research
Research in Science and Technology Studies (STS) has begun to probe the limitations of representing bodily knowledge through computing. In her ethnographic examination of surgical simulation, STS scholar Rachel Prentice [12] describes the process of ‘mutual articulation’ or the shaping of bodies by machines and machines by bodies. Both the bodies of the surgeons and those of the patients on which they practiced became present and knowable through the simulator. Building on Latour’s [11] notion of ‘articulation’ through which bodies and knowledge get mutually enacted, she writes:

[...] researchers pooled various disciplinary knowledges of anatomy, surgery, computation, education, cognition and engineering to develop an object (a model, a software program, a device) that has a particular relationship to the user’s body. At each point, then, researchers are working to create interpretations of what human bodies are in relation to these objects; that is, to articulate the body in new ways [12:862].

Body objects, according to Prentice, emerge as representations of human bodies that inhabit computers for use by people and computers. Drawing
on this category, our design team will use Synku to investigate what it would mean to build sensory objects, or algorithmic representations of sensory experiences for interpretation and use by people and software. For interaction design, sensory objects exemplify a type of 'strong concept' [9], an intermediate level of knowledge between theories and particular instances of design. Given that self-tracking activities have become increasingly shared [2], we wonder what form such sensory objects may take within everyday contexts of communication. How might we use sensory objects to articulate the body in new ways? What are the consequences of these new forms of articulation for understanding and engagement? We use design research to speculate on these questions: exploring and intermingling how people see through the eyes of a vision algorithm and hear through the ears of sound processing.

In the months ahead, we plan to expand our study of SynKu in two ways. Empirically, we plan to launch a study of our application within a local art community in Seattle, Washington in partnership with Seattle's Design Festival "Design in Public," a two-week affair that features public demonstration and celebration of the Seattle design community. Theoretically, we will use the notion of sensory objects - or the processing, interpretation and use of sensory experience - to investigate new forms of articulation in HCI.

References