GIERAD LAPUT | RESEARCH STATEMENT

My research in **human-computer interaction** designs, builds, and evaluates novel interactive technologies that greatly enhances **input expressivity for users** and **contextual awareness for devices**. My work lies at the intersection of interactive systems, sensing, and applied machine learning. I have published over 25 papers in this space, with several more in review, six of which have received Best Paper Awards and Nominations at premiere venues in human-computer interaction.

INTRODUCTION

When people converse, **explicit** (*e.g.*, speech, pointing) and **implicit** cues (*e.g.*, gaze, body language, facial expressions) play key roles in creating nuanced and expressive communication. We can draw similar parallels for interactive systems. Although computers have rich output capabilities (*e.g.*, screens, sound, haptics), they lack input richness. Most systems are driven by a limited set of **explicit input** modalities (*e.g.*, typing, touch, voice). For example, mobile and wearable computers feature small screens, and smart speakers only offer linear voice commands that result in suboptimal expressiveness. A main thread of my Ph.D. research developed new ways to expand explicit input richness across different modalities [6, 7, 8, 9, 10, 13], making key contributions in this area.

Furthermore, systems lack **implicit inputs** and situational awareness to foster nuanced and assistive human-computer interactions. Most interactive systems have **no implicit input** channel, primarily due to lack of contextual sensing and understanding. For example, wearables and Internet-of-Things (IoT) devices are oblivious to their physical context, despite being right there, strapped on a user's arm or sitting on a kitchen countertop – a missed opportunity. As new platforms proliferate across different facets of our everyday lives, increasing implicit input bandwidth will become highly important if we are to power more human-centric experiences. Digitizing the physical environment through context-awareness has many high-impact applications, from specific domains such as elder care, health monitoring, and empowering people with disabilities, to much broader applications such as smart infrastructures, robotics, and novel interactive experiences for consumers.

THESIS: CONTEXT-DRIVEN IMPLICIT INTERACTIONS

Most approaches for imbuing computers with contextual awareness and sensing are impractical – often costly, requiring instrumentation of environments, and typically special-purpose, limiting the range of activities and objects systems can sense. In addition to these practical constraints, equally challenging is that raw sensor output rarely matches human intention and semantics. For example, an accelerometer provides coarse movement information, but rarely fine-grained human activity. Similarly, a door sensor may fall short in answering a user's true question, *e.g.*, "*are my kids home from school?"* These questions are difficult to answer by the conventional sensors we deploy today. Therefore, the "holy grail" for achieving the promise of nuanced, highly ubiquitous, context-aware systems rests on exploring sensing modalities that are low-cost, and easy to deploy. Likewise, such sensing systems should support "virtualization" of low-level signals into semantically high-level, user-relevant information.

My Ph.D. thesis focused on the construction and evaluation of sensing technologies that can be practically deployed, and yet still greatly enhance contextual awareness, primarily drawing upon machine learning to unlock a wide range of applications. I attack this problem area on two fronts: 1) supporting sensing expressiveness via context-sensitive wearable devices, and 2) achieving general-purpose sensing through sparse environment instrumentation. I built algorithms that extract meaningful signals and patterns from sensor data to enable high-level abstraction and interaction. In my Ph.D., I have established a body of work in this area, which have received multiple awards, spun-off as startups, and licensed by Fortune 500 companies for commercialization. An array of follow-up work is in review [11, 12, 13], building on the research discussed here.

The Arm as a Context-Sensing Springboard

In the past few years, wearables have emerged that afford a unique, on-body beachhead for implicitly sensing human activity. Prior activity sensing work has chiefly focused on detecting human locomotion, such as walking, running, sleeping and driving. While whole-body activities are clearly important, it would also be useful to detect fine-grained actions, such as what activities the hands are engaged in. As philosopher Immanuel Kant argued, *"The hand is the visible part of the brain."* Indeed, such fine-grained detection could serve to make computational experiences much more context-sensitive. This offers great potential for transforming the user's arms and hands into an expressive context-sensing platform. Instead of smartwatches being viewed as a handicapped versions of smartphones, users should instead view smartwatches as an opportunity to unlock version 2.0 of their arms.

As people use their hands, micro-vibrations propagate into the arm, carrying information about the objects they interact with and the activities they perform throughout the day. Smartwatches are ideally situated to capture these vibrations, as I showed In **ViBand** (https://youtu.be/Poi0MeASmuY) [1]. More specifically, I overclocked a smartwatch accelerometer into a rarely-used, high-speed sampling mode: 4000 Hz. At this high sample rate, ViBand can capture small compressive waves (*i.e.*, bio-acoustics) propagating through the user's arm (Figure 1). Unlike microphones, bio-acoustics are physically coupled to the body, making this technique naturally resistant to external noise. By combining this novel signal source with supervised classification models, ViBand unlocks a wide range of applications. For example, ViBand can appropriate the arm into an input surface by localizing on-body taps, and classifying rich hand gestures such as scratches, flicks, and claps. It can also classify fine-grained hand actions such as chopping, hand washing, and toothbrushing. Objects such as hand saws, coffee grinders, and vacuum cleaners generate vibrations that propagate through the hands, which ViBand can detect and automatically launch context-relevant functionality. By being software-only, ViBand could be deployed to existing smartwatches with an over-the-air update.

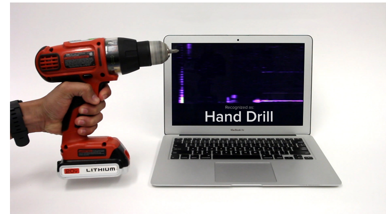

Figure 1. ViBand (worn) overclocked a wearable accelerometer to 4000Hz, enabling the detaction of object interactions from bio-acoustic signals.

Of course, sensing bio-acoustics is insufficient by itself since many objects and events do not generate physical vibrations. In response, I sought to expand the range of signals that can be sensed from the hand. In **EM-Sense** (https://youtu.be/fpKDNle6ia4) [2], I leveraged the conductive property of the human body, and built a wrist-worn device that detects the electro-magnetic (EM) noise emitted by many everyday electrical and electromechanical objects, such as kitchen appliances, computing devices, power tools, and automobiles. These signals tend to be highly characteristic to these objects, owing to their unique internal operations (*e.g.*, brushless motors, capacitive touchscreens) and different enclosure designs, material composition, and shielding. When a user makes physical contact with these objects, EM signals propagate through the body and can be sensed from the watch. EM-Sense detects and classifies these signals in real time, enabling robust, on-touch object detection in a compact form factor. By learning EM "signatures", EM-Sense can discriminate between scores of objects, independent of wearer, time, and environment. In one of several studies, EM-Sense achieved a mean accuracy of 96.1% across dozens of objects, multiple users, multiple locations, and using data trained on a single independent user collected six weeks prior. This capability, combined with ViBand, could power many context-based applications, from passive monitoring of wearer's daily routines, to proactive suggestions (*e.g.*, recipes and routes) based on user-object interactions.

General-Purpose Ubiquitous Sensing

ViBand and EM-Sense unlock promising opportunities for wearables as context-sensing springboards, but user instrumentation can be a high bar for adoption. For example, kitchens, classrooms, and even nursing homes could benefit from contextual sensing, but it is unrealistic to expect all users in that space to wear a wrist-worn device. Therefore, a complementary approach to wearable sensing is to embed sensors at key probe points in the environment. This is an active research area, and I created a taxonomy that represents prior work in this space (Figure 2). Along the y-axis is the number of distinct sensed facets (*e.g.*, states and events), while the x-axis is the number of sensors needed to achieve this output. A single room can have dozens of complex environmental facets worth sensing, but the cost of physical sensors is significant, not including the even greater cost of deployment and maintenance. Therefore, the ideal sensing approach occupies the top-left of Figure 2, wherein *one* sensor can enable *many* sensed facets, and more specifically, beyond any one single instrumented object. This general-purpose sensing approach is challenging, as it must be inherently indirect to achieve this breadth, but it has the potential to transform spaces into *smart environments* in a practical way.

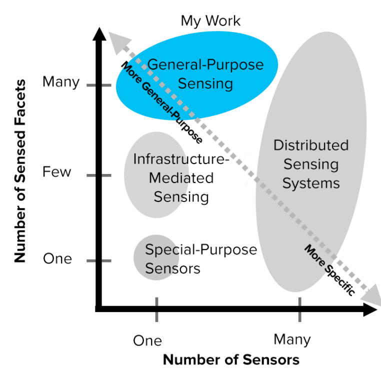

Figure 2. A taxonomy illustrating key approaches in environmental sensing. My work highlighted in blue.

Computer vision (CV) has come closest to achieving the goal of general-purpose sensing because cameras offer rich, indirect data that can be processed through machine learning to yield sensor-like feeds. In **Zensors** (https://youtu.be/VVP9emuFsQI) [3], I repurposed disused mobile devices and their cameras into "universal" sensors. I fused supervised classification models with crowdsourcing to provide instant, human-intelligent environmental sensors that users can set up without specialized training. This hybrid crowd-AI model allows a "sensor" to be automated over time, enabling questions such as "is a parking spot occupied?", "is the basement flooded?" or "is the trash full?" By exposing this functionality as APIs, Zensors enables rich data abstractions for context-aware environments.

While Zensors, and CV-based sensing approaches in general are powerful, cameras have been widely studied and recognized for their high level of privacy invasion and social intrusiveness, and thus carry a deployment stigma. In **Synthetic Sensors** (https://youtu.be/aqbKrrru2co) [4], I achieved much of the same sensing versatility and accuracy as Zensors, but using only low-level, non-camera sensors. The flexibility of Synthetic Sensors is driven by a novel "sensor tag" I created (Figure 3), equipped with a suite of denatured low-level sensors that indirectly detect events manifested in an environment. For example, when a faucet is turned on, a nearby sensor tag can pick up the vibrations induced by service pipes installed behind the wall, as well as the characteristic wide-band acoustic features of running water. Using an ensemble of classifiers, Synthetic Sensors virtualize sensor data into actionable feeds, powering end-user applications. When plugged into a kitchen wall, events such as the microwave running, stovetop burner in use, or a blender in operation can be detected. Likewise, when installed in a bathroom, the system can detect events such as the shower running, toilet flushing, or a hair dryer in use, all from a single sensor. In one of many studies, I show that the system achieves a mean sensing accuracy of 96%, across 40 physical events in multiple locations, spanning several weeks.

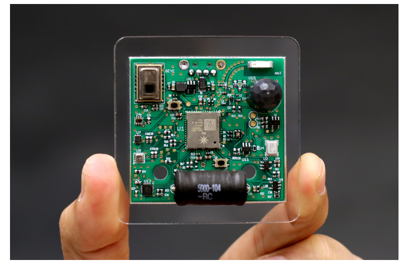

Figure 3. I built a sensor tag (plugged-in through a wall socket), comprised of nine sensors capturing twelve unique sensing dimensions.

Synthetic Sensors offers a practical vehicle for general-purpose sensing, but it fundamentally incurs a heavy data collection and training burden. To build robust models, large amounts of data is required, which is time consuming, error-prone, and difficult to scale (especially if end users are involved). In **Ubicoustics** (https://youtu.be/N5ZaBeB07u4) [5], I address this problem by leveraging microphones for activity sensing. Microphones are one of the most common sensors found in consumer devices but weakly utilized as a contextual sensing platform. Further, microphones are the exact sensors used to record *professional sound effect libraries* traditionally used in the entertainment industry. Unlike internet-mined datasets (which are noisy and weakly labelled), professional sound effects contain noise-free, well-segmented, well-labeled, and highly diverse everyday sounds. Because of its atomic properties, they can be easily transformed and projected into hundreds of realistic environments (*e.g.*, applying impulse functions and mixing with background tracks), synthetically growing a dataset with a high signal-to-noise ratio. Armed with a corpora of context-specific, high-quality data, I fine-tune existing state-of-the-art deep learning models (*e.g.*, leveraging pre-trained weights), enabling sound activity recognition without *in-situ* training. My evaluations, conducted across 30 activities, 50 locations and multiple device platforms (*e.g.*, from watches, to smart speakers, and IoT sensors), show that models tuned with Ubicoustics can achieve superior accuracy than those trained on internet-mined data alone. More surprising is that system accuracy is comparable to human-level performance (baselined on 600 human coders). By leveraging sensors and high-quality data that already exist, Ubicoustics offers another promising path towards practical context sensing.

FUTURE RESEARCH AGENDA

In the future, I plan to continue my line of work in enhancing input expressivity and contextual awareness in human-computer interactions, but also expanding to other domains (*e.g.*, sensing for robotics)**.** As faculty building a research group, I am also extremely excited to explore related areas of interest. For example, I want to scale-up in physical scope – my work has traditionally explored room-scale domains, but I want to expand this to **buildings, neighborhoods, and cities**. This is an inherently interdisciplinary endeavor, and I look forward to bringing a new set of tools for researchers and practitioners in areas such as **architecture** and **urban planning**. Additionally, as sensors and devices proliferate, I want to explore how to mitigate and balance their effect on **privacy.** My training in human-computer interaction will allow me to tackle this problem from both social and computational perspectives. This includes working with social scientists to create better privacy models that limit impedance mismatch between data fidelity and user trust. Equally important is working with system designers and machine learning experts to relentlessly **push AI more closer to the "edge,"** and build better **differential privacy** models. Finally, I am also excited to explore a broader view of sensing, for example, investigating better approaches to support **sensor data visualization** and **sensemaking.** As previously mentioned, encoding sensor data with human-level semantics will play a key role in unlocking context-driven applications, and I want to design, build, and study systems that empower users to make better sense of large volumes of sensor data.

REFERENCES

1. **Gierad Laput**, Robert Xiao, and Chris Harrison. 2016. ViBand: High-Fidelity Bio-Acoustic Sensing Using Commodity Smartwatch Accelerometers. In Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16).
2. **Gierad Laput,** Chouchang Yang, Robert Xiao, Alanson Sample, and Chris Harrison. 2015. EM-Sense: Touch Recognition of Uninstrumented, Electrical and Electromechanical Objects. In Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15).
3. **Gierad Laput,** Walter S. Lasecki, Jason Wiese, Robert Xiao, Jeffrey P. Bigham, and Chris Harrison. 2015. Zensors: Adaptive, Rapidly Deployable, Human-Intelligent Sensor Feeds. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15).
4. **Gierad Laput,** Yang Zhang, and Chris Harrison. 2017. Synthetic Sensors: Towards General-Purpose Sensing. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17).
5. **Gierad Laput,** Karan Ahuja, Mayank Goel, and Chris Harrison. 2018. Ubicoustics: Plug-and-Play Acoustic Activity Recognition. In The 31st Annual ACM Symposium on User Interface Software and Technology (UIST '18).
6. **Gierad Laput**, Xiang 'Anthony' Chen, and Chris Harrison. 2016. SweepSense: Ad Hoc Configuration Sensing Using Reflected Swept-Frequency Ultrasonics. In Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI '16).
7. **Gierad Laput**, Eric Brockmeyer, Scott E. Hudson, and Chris Harrison. 2015. Acoustruments: Passive, Acoustically-Driven, Interactive Controls for Handheld Devices. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15).
8. **Gierad Laput**, Eytan Adar, Mira Dontcheva, and Wilmot Li. 2012. Tutorial-based interfaces for cloud-enabled applications. In Proceedings of the 25th annual ACM symposium on User interface software and technology (UIST '12). ACM, New York, NY, USA, 113-122.
9. **Gierad Laput**, Mira Dontcheva, Gregg Wilensky, Walter Chang, Aseem Agarwala, Jason Linder, and Eytan Adar. 2013. PixelTone: a multimodal interface for image editing. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13). ACM, New York, NY, USA, 2185-2194.
10. **Gierad Laput**, Robert Xiao, Xiang 'Anthony' Chen, Scott E. Hudson, and Chris Harrison. 2014. Skin buttons: cheap, small, low-powered and clickable fixed-icon laser projectors. In Proceedings of the 27th annual ACM symposium on User interface software and technology (UIST '14). ACM, New York, NY, USA, 389-394.
11. **Gierad Laput**, and Chris Harrison. 2019. Exploring the Efficacy of Sparse, General-Purpose Sensor Constellations for Wide-Area Ubiquitous Sensing. *Under Review*. In Proceedings of the ACM Journal on Interactive, Mobile, Wearable, and Ubiquitous Technologies (IMWUT UbiComp '19).
12. **Gierad Laput**, and Chris Harrison. 2019. Investigating Fine-Grained Hand Activities Using Commodity Smartwatches. To appear in Proceedings of the 37th Annual ACM Conference on Human Factors in Computing Systems (CHI '19).
13. **Gierad Laput**, and Chris Harrison. 2019. SurfaceSight: A New Spin on Touch, User, and Object Sensing for IoT Experiences. To appear in Proceedings of the 37th Annual ACM Conference on Human Factors in Computing Systems (CHI '19).