SweepSense: Ad Hoc Configuration Sensing Using Reflected Swept-Frequency Ultrasonics

Gierad Laput   Xiang 'Anthony' Chen   Chris Harrison
Human-Computer Interaction Institute
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15217
{gierad.laput, xiangche, chris.harrison}@cs.cmu.edu

ABSTRACT

Devices can be made more intelligent if they have the ability to sense their surroundings and physical configuration. However, adding extra, special purpose sensors increases size, price and build complexity. Instead, we use speakers and microphones already present in a wide variety of devices to open new sensing opportunities. Our technique sweeps through a range of inaudible frequencies and measures the intensity of reflected sound to deduce information about the immediate environment, chiefly the materials and geometry of proximate surfaces. We offer several example uses, two of which we implemented as self-contained demos, and conclude with an evaluation that quantifies their performance and demonstrates high accuracy.

Author Keywords: Acoustic sensing; mobile devices; interaction techniques; novel input.

ACM Classification: H.5.2 [Information interfaces and presentation]: User Interfaces - Input devices and strategies.

INTRODUCTION

Today, almost all general-purpose computing devices contain a speaker and microphone. We propose utilizing these ubiquitous sensors to bring novel sensing abilities to devices without extra or special hardware (e.g., [8,11]). Specifically, we emit inaudible frequency sweeps using the built in speaker, and capture the reflected waveforms using the built in microphone. These sounds are then analyzed and classified to infer some aspect of the environment.

Our work was originally inspired by the field of soundscape ecology [18]. We learned that different environs (natural or otherwise) have significantly different acoustic properties—the composition and spatial configuration of flora and geological features act as acoustic filters, passing some sounds while attenuating others. This can introduce selective pressure, leading to e.g., animals communicating using frequencies that carry farthest in their habitats (see e.g., [5,12]).

This inspired us to run experiments in our lab, sweeping through a range of frequencies emitted from a laptop, and looking at attenuation due to the environment. We noticed that changes in physical configuration could alter the signal, for example, a door being opened or the lid of our laptop being repositioned; it can even capture subtle relationships between wearable devices and the human body, such as knowing when headphone buds are removed from the ears.

Such actions appear to expose different facets and/or materials, with different acoustic reflectance characteristics. Additionally, in enclosed "chambers", such as the ear canal, cars, rooms and similar, certain frequencies may become standing waves [6], with amplifying effects. Finally, a sound wave reflected off multiple surfaces will return to the microphone at different phases, producing characteristic interference effects. These effects all interact to produce a "sound signature", which can be used for classification.

RELATED WORK

Modal analysis [2] and swept frequency acoustic interferometry (SFAI) are well known in fields such as material identification [21], structural analysis [7] and petrogeology [4]. In the HCI domain, acoustic-based input sensing can be categorized into three general approaches. First are time difference of arrival (TDoA) and time of flight (ToF) localization techniques, most commonly in the form of "sonar," which has seen extensive use (see e.g., [19,23]). Additionally, sensing acoustic Doppler shifts [1] has also proven to be useful (see also [24] for an overview).

More similar to our approach are acoustic fingerprint-based techniques, which use an uncontrolled, but characteristic signal to distinguish between possible states. Approaches can be active—as in the case of Mujibiya et al. [15] and Takemura et al. [22]–or passive, such as Skinput [10]. Conceptually related to our approach is Touch & Activate [16], which enabled ad-hoc touch sensing on objects using swept frequency vibrations. In contrast to our technique, the latter used a piezo contact microphone and a piezo buzzer physically attached to an object and relied chiefly on physical resonance (as opposed to reflectance). The authors cite Touché [20] as inspiration, which used swept frequency *electrical* signals for ad hoc capacitive sensing. As discussed in both papers, applications can often suffice without a full sweep. However, this requires *a priori* knowledge of

discriminative frequency bands in different contexts, which precludes general and immediate use.

SWEEPSENSE

Our approach, which we call *SweepSense,* can open new interactive opportunities in contexts where there are speakers and microphones present. Using a device's built-in speaker, we emit a repeating 20ms ultrasound linear frequency sweep. Acoustic sweeps can be generated programmatically (e.g., using various audio APIs such as minim, and BASS audio I/O) or pre-cached using synthesis software (e.g., Audition, Audacity). Further, we use low-level audio API calls (e.g., using callbacks whenever audio samples are captured) to more closely monitor incoming data from the microphone. Most audio APIs expose these types of callbacks, such as the `sample(float[])` function call in minim, or the `setRenderCallback()` in iOS CoreAudio. Meanwhile, captured audio samples are stored in a circular buffer (~100ms or less), where frequency analysis and machine learning are eventually performed.

Next, the captured signals are transformed into frequency space using a Fast Fourier Transform (FFT) and only the inaudible spectrum components of interest are saved (e.g., 20-40kHz for a device capable of sampling at 96kHz). In addition to the raw FFT values, we also compute a series of standard features (derived from [17]): RMS, average power, spectral center of mass, max/min index and values, standard deviation, and spectral band ratios. These features are passed to a Sequential Minimal Optimization-based Support Vector Machine (SMO-SVM), provided by the Weka Toolkit [9]. Note that this model must be trained with data before it can provide real-time classification.

EXAMPLE IMPLEMENTATIONS

We implemented two illustrative applications as self-contained demos. These represent different scales of use, and serve as examples of discrete and continuous sensing.

Discrete Sensing: Smart Ear Buds

SweepSense may be useful in small-scale applications, such as ear buds (Figure 1). For example, when both ear buds are in, music plays as usual. However, if both ear buds are removed, the music can be automatically paused. Also, a phone call could be answered by removing a single bud.

Implementation. For our ear bud application, we used a pair of generic, Samsung in-ear headphones. To avoid crosstalk, we emit a 20 to 22kHz sweep through the left ear bud and a
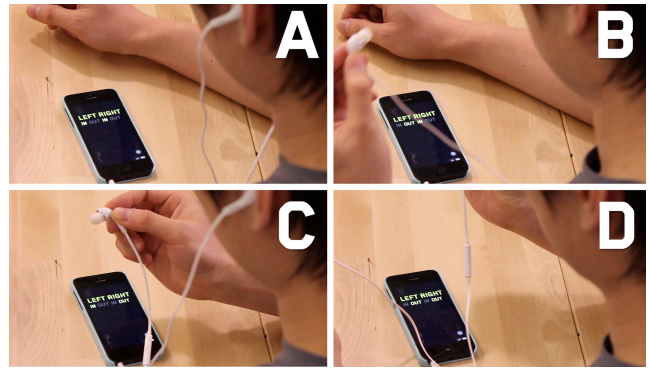

Figure 1. SweepSense can detect whether (A) both ear buds are in, (B) left is out, (C) right is out, or (D) both are out.

23 to 25kHz sweep through the right bud. These headphones contain an integrated microphone approximately 20cm down the cord, which we use to capture sound. Using this setup, we can detect four possible states: 1) both buds in, 2) left bud out, 3) right bud out, and 4) both buds out (Figure 1 and Video Figure). We used a SVM model trained using Weka's Quad SMO [9] with default parameters.

Other researchers have achieved similar functionality, but only through additional hardware. For example, small capacitive sensors can be added to ear buds to detect skin contact, as demonstrated in Buil et al. [3]. Metzger et al. used infrared proximity sensors [14] to detect gestures performed around the ear. Most recently, Manabe and Fukumoto [13] showed that a low cost supplementary circuit could be used to detect physical taps on headphones, which physically actuate the speaker diaphragm. SweepSense achieves equivalent functionality and could be integrated to existing devices with just a simple software update.

Continuous Sensing: Laptop Lid Configuration

Contemporary laptops use a special purpose sensor to detect lid configuration, but this is presently limited to detecting open or closed states (i.e., binary). SweepSense could be used to infer *continuous* angular position of the lid without any hardware modification (Figure 2). This could be used, for example, to lock the screen when the lid is lowered, or to sense lid-based gestures e.g., leaning backward to trigger full-screen, or wiggling to minimize windows.

Implementation. For our laptop lid angle application, we emit a continuous 20-40kHz sweep from a 2013 MacBook Air's 2013 built-in speakers (located where the display

Figure 2. SweepSense can infer continuous laptop lid angle. Here the lid is positioned at 100, 70, 50 and 30° (A-D); Sweepsense reports 97, 70, 51 and 30° respectively. Far right (E), a 1Hz "wiggle" gesture is performed and visible in the raw signal.

meets the main chassis). The microphone is located on the left side bezel of the base. Our machine learning setup uses an SVM regression model with a radial basis function kernel (γ=0.01), trained using Weka's SMOReg implementation [9] using default optimization parameters. Our regression model was trained to infer angular positions between 30° and 110°, allowing our system to behave like a virtual lid angle sensor (see Fig. 2 and Video).

We note that although lid sensing can be achieved using accelerometers, it is rare for laptop lids to have accelerometers solely for this purpose. Meanwhile, most laptops have a speaker and microphone. Thus, SweepSense brings new capabilities through software, and more critically, without any additional hardware.

USER STUDY

We sought to evaluate the feasibility of our approach through multiple user studies. In response, we recruited two separate groups of participants, (32 total, 11 female, mean age=24, STDEV=5.5), and each experiment took about 30 minutes. All participants were paid $10 for their time. This section describes the studies in detail.

Ear Bud Experiment

To evaluate ear bud configuration sensing, we had 24 participants (9 female, mean age=24) use our off-the-shelf, Samsung-branded in-ear headphones in two different locations: a large, open social area in an academic building and a medium-sized office (12 participants for each location). Although we created an iOS version of our SweepSense recognizer (Figure 1 and Video Figure), we used a laptop to run the experiment. To better emulate real-world use, music was played in the ear buds throughout the experiment.

Following a brief explanation, participants were asked to replicate a series of ear bud configurations (depicted in Figure 1), which were shown as photographs on the laptop screen. Each configuration was requested five times in a random order (during which time ten classifications were made over a ten second period). Of note, the system performed live classification with no per user training or calibration. In total, we collected 1200 classification attempts for each configuration. Overall accuracy was 94.8% (SD=1.3%); Table 1 provides a confusion matrix.

Laptop Lid Angle Experiment

For the ear bud experiment, it was important to capture test data from a variety of users, as everyone's ears are different. However, since the laptop was constant in this experiment, we recruited a smaller pool of eight participants (2 female, mean age=25). Further, we wanted to test whether SweepSense could generalize across a *range* of environments. For this experiment, we used a classifier trained specifically for our MacBook Air 2013. Meanwhile, we visited participants in a location of their choosing, which included a variety of commonplace settings such as a cafe, study hall, food court, conference room, and shared lab space (mean ambient noise 55.6dB, SD=5.72dB). Overall, we found no significant effect based on location.

	Both-in	Left-out	Right-out	Both-out	Accuracy
Both-in	1150	15	28	7	*95.8%*
Left-out	70	1112	1	22	*92.7%*
Right-out	5	0	1142	53	*95.2%*
Both-out	1	0	51	1148	*95.7%*

Table 1. Confusion matrix for earbud placement classification.

In this experiment, a small *protractor* was placed along the right side of the laptop, near the hinge, allowing participants to accurately orient the lid to any requested angle. A randomized list of angles from 30° to 110° in 10 degree increments was requested verbally (e.g., "please set the lid to 60 degrees"). Each angle appeared twice, for a total of 18 trials. Once the participant was satisfied with their lid positioning, 20 angular estimates (captured over a period of approximately three seconds) were made using the pretrained classifier and averaged to produce a final estimate.

Results show that the mean angular error was ±4.1° (SD=1.7°) for angles between 30°-100°, summarized in Figure 4. Note that error increases substantially beyond 100°, which we hypothesize is due to the screen no longer strongly participating in reflecting sound downwards towards the microphone (i.e., sound mostly emanates away from the laptop). Finally, our regression model achieved a correlation coefficient (R^2) of 0.988, evaluated using 10% of the dataset we withheld for testing.

DISCUSSION

There are several drawbacks of note. One is our dependence on low frequency ultrasound, which may be audible to children, some adults and pets. However, the technique is not reliant on any single frequency, and our machine learning approach makes it easy to increase the lower bound frequency assuming the hardware supports it. In practice, this audibility issue does not appear to be a critical problem given that ultrasonic sensors have enjoyed commercial success for decades (e.g., automatic door openers) without issue.

Secondly, we found that many devices do not support high sampling rate audio output (we utilized 96kHz for our laptop example, but found that 44.1kHz is a more common upper bound). This reduces audio output fidelity, especially in the ultrasonic range. However, systems today are not engineered for this purpose, and so with purposeful integration, higher ultrasound frequencies might be better supported for applications like we suggest in the future.
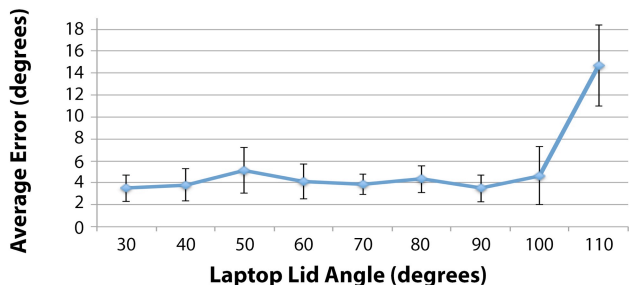
Figure 4. Our regression model can predict laptop lid angle. Below 100°, the mean angular error is ±4.1°.

Finally, acoustic leakage is another issue. For example, if ear buds are partially inserted, our classifiers will struggle to make accurate predictions. More training data would help, but as with other classification systems, this may be a significant challenge. Finally, classifiers may need to be device and model specific—for example, our MacBook lid angle demo is not immediately portable to a Dell laptop.

In general, SweepSense introduces interactive opportunities in environments where speakers and microphones are present. One possibility for future work is in living rooms, which often contain speakers for entertainment and where microphones are increasingly common (e.g., Microsoft Kinect, Smart TVs). This could allow, for example, occupant sensing and detecting whether people are sitting or standing, which could be used to e.g., intelligently pause media. Meanwhile, vehicles generally contain speaker systems, and increasingly feature microphones for Bluetooth-connected telephony and voice navigation functions. SweepSense could allow for occupancy sensing in all seats, enabling e.g., richer safety applications. Few cars today track the status of their windows, and typically only know if doors are opened or closed. With SweepSense, cars could infer the *analog* state of both windows and doors (albeit susceptible to noise in certain cases).

CONCLUSION

Uniquely, and in contrast to prior work, SweepSense uses in-air, reflected, swept-frequency ultrasound. This approach adds to a growing body of work that can take advantage of devices' existing speakers and microphones, allowing such techniques to be deployed with a simple software update. In the future, we hope to explore larger contexts that have speakers and microphones, such as public address (PA) systems in subway stations and supermarkets, and sound systems at concert venues and stadiums.

REFERENCES

1. Aumi, M., Gupta, S., Goel, M., Larson, E. and Patel, S. DopLink: using the doppler effect for multi-device interaction. In *Proc. UbiComp '13*. 583-586.

2. Brandt, A. (2011). "Noise and Vibration Analysis." John Wiley & Sons, Ltd., UK.

3. Buil, V., Hollemans, G., and Wijdeven, S. van de. Headphones with touch control. *Proc. MobileHCI '05*. 377-378.

4. Chevron Corporation. "Chevron, GE Form Technology Alliance." Retrieved April 11, 2014 from: http://investor.chevron.com/phoenix.zhtml?c=130102&p=irol-newsArticle_Print&ID=1895989

5. Cocroft, R and Rodríguez, R. The Behavioral Ecology of Insect Vibrational Comm. *BioSci.*,55. '05. 323-334.

6. Cox, T., D'Antonio, P. and Avis, M. Room sizing and optimization at low frequencies. *Journal of the Audio Engineering Society*, 52(6), 2004, pp. 640-651.

7. Gloth, G. and Sinapius, M. Analysis of swept-sine runs during modal identification. *Mechanical Systems and Signal Processing*, 18(6), Nov. 2004, pp. 1421-1441

8. Gupta, S., Morris, D., Patel, S. and Tan, D. SoundWave: using the doppler effect to sense gestures. In *Proc. CHI '12*. 1911-1914.

9. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. The WEKA Data Mining Software: An Update; *SIGKDD Explorations*, 11(1), 2009.

10. Harrison, C., Tan, D. and Morris, D. Skinput: appropriating the body as an input surface. In *Proc. CHI '10*. 453-462.

11. Hwang, S., Bianchi, A. and Wohn, K. VibPress: estimating pressure input using vibration absorption on mobile devices. In *Proc. MobileHCI '13*. 31-34.

12. Kime, N., Turner, W. and Ryan, M. The transmission of advertisement calls in Central American frogs. *Behavioral Ecology*, 11(1), 2000, pp. 71–83.

13. Manabe H. and Fukumoto, M. Headphone taps: a simple technique to add input function to regular headphones. In *Proc. MobileHCI '12*. 177-180.

14. Metzger, C., Anderson, M., Starner, T. FreeDigiter: A Contact-free Device for Gesture Control. *Proc. ISWC '04*. 18-21.

15. Mujibiya, A., Cao, X., Tan, D., Morris, D., Patel, S. and Rekimoto, J. The sound of touch: on-body touch and gesture sensing based on transdermal ultrasound propagation. In *Proc. ITS '13*. 189-198.

16. Ono, M., Shizuki, B. and Tanaka, J. Touch & activate: adding interactivity to existing objects using active acoustic sensing. *In Proc. UIST '13*. 31-40.

17. Peeters, G. A large set of audio features for sound description similarity and classification in the CUIDADO project. Tech. report, ICRAM, 2004.

18. Pijanowski, C., Villanueva-Rivera, J., Dumyahn, L., Farina, A., Krause, L., Napoletano, M., Gage, H. and Pieretti, N. Soundscape Ecology: The Science of Sound in the Landscape. *BioScience*, 61(3), 2001, pp. 203.

19. Priyantha, N. B., Chakraborty, A., and Balakrishnan, H. The Cricket Location-Support system. In *Proc. MobiCom '00*. 32-43.

20. Sato, M., Poupyrev, I. and Harrison, C., Touché: enhancing touch interaction on humans, screens, liquids, and everyday objects. In *Proc. CHI '12*. 483-492.

21. Sinha, D., Springer, K., Han, W., Lizon, D., and Houlton, R. Swept-frequency acoustic interferometry technique for noninvasive chemical diagnostics. In *Proc. 3rd International Conference On-Site Analysis*, 1995.

22. Takemura, K., Ito, A., Takamatsu, J. and Ogasawara, T. Active bone-conducted sound sensing for wearable interfaces. In *Proc. UIST '11 Adjunct*. 53-54.

23. Tarzia, S., Dick, R., Dinda, P. and Memik, G. Sonarbased measurement of user presence and attention. In *Proc. Ubicomp '09*. 89-92.

24. Watanabe, H., Terada, T. and Tsukamoto, M. Ultrasound-based movement sensing, gesture-, and context-recognition. In *Proc. ISWC '13*. 57-64.