UWHear: Through-wall Extraction and Separation of Audio Vibration Using Wireless Signals

Ziqi Wang, Zhe Chen, Akash Deep Singh, Luis Garcia, Jun Luo, Mani B. Srivastava

The Problem Definition:

The detection and classification of acoustic events is a booming area of research and is used/required in a lot of smart systems to make rich inferences like understanding human behavior, effective smart home devices, etc. Traditionally, this is done using microphones as sensors or using RF (which doesn't support the separation of sound from multiple sources). The paper very well tries to fill this gap and presents the idea of the UWHear, as a system that will recover and separate sounds from multiple sources simultaneously.

Key Idea:

The key idea is to construct a fine-grained audio sensing system using the Impulse Radio Ultra-Wideband (IR-UWB) technology to recover and separate sound from multiple sources. The expectations are that the system will be robust in non-line-of-sight (NLOS) scenarios. For every transmitted probe pulse, the IR-UWB receiver may collect several reflected pulses. Consequently, sound sources can be well separated by accurately estimating the Time-of-Flight (ToF) of the reflected pulses. This procedure is repeated rapidly with constant intervals to produce two-dimensional data. UWHear uses a commercial-of-the-shelf (COTS) IR-UWB radar sensor with optimal driver settings, as well as a pure statistical signal processing pipeline. The system is implemented with Novelda Xethru X4M05 IR-UWB radar board combined with a Raspberry Pi 3B+. The max. range detection of the solution is 8m when the device is in direct LOS and if the device & the sensor are separated by a light material wall then the system can operate within the range of 2.5m.

Important Details:

1. IR-UWB radars send short pulses in the time domain while occupying a wide frequency bandwidth.
2. UWHear is capable of through-wall sensing of audio vibrations.
3. The data structure of UWHear consists of frames & distance bins. The data collected from the departure of the probe pulse to the arrival of the last response is called a frame. The Y-axis denotes the slow time and X-axis denotes the fast time. The fast time denotes the round trip ToF of a pulse, hence we can convert the fast time into distance bins.
4. The theoretical analysis in the paper proves that the sound waveform values are proportional to the amplitude of the in-phase or quadrature part of the filtered sliced data.
5. The fast time is fine-grained in tens of picoseconds, while slow time has a scale of hundreds of microseconds.
6. The amount of target micro displacement is linearly proportional to the amplitude of the quadratic part of the receiving signal.

7. Sound-related vibration information can be extracted by analyzing the amplitude of the in-phase or quadrature of the UWB receiving signal, whichever gives a higher signal-to-noise ratio.
8. Phase Noise Correction Algorithm is used to remove sampling clock jitters, Static Clutter Suppression suppresses the reflections caused by static objects, and Vibration Activity Localization determines the distances of the vibrating targets. Further, the Spectral subtraction algorithm is used for denoising.
9. The maximum range of the IR-UWB radar can be as far as 9.87m.
10. Due to the "crosstalks" between the transmitting antenna (tx) and the receiving antenna (rx), i.e., direct signal leakage from the tx to the rx, the signal amplitude in the first few distance bins is always high. This is leveraged for a baseline for phase calibration.

<u>My Thoughts and Criticism:</u>

The paper indeed presents some great work in the field with relevant observations and theoretical proofs. The paper provides an unbiased view of the critics, loopholes, and future work, which demonstrates a great mindset. It can be observed very clearly that the author has tried to be open about his ideas and thought process on the factors ignored, and possible counters/critics of the work done.

However, a few of the assumptions and scenarios that led me to question, and think are:

1. Vibrations and transmission will also be impacted by the medium of transmission as well as the airflow between the sensor and transmitter. This factor should be tested extensively.
2. Currently, the prototype is designed in a way that it will fail to promise accuracy outdoors due to increasing SNR.
3. The author has used a slow-time window of 1s to perform vibrating target localization, which I believe might not fit objects of varied sizes as well as vibration intensity.
4. There are certain observations made by the authors while testing their solution which they have considered as "probably" due to the electromagnetic field of the circuit, etc. But nothing is mentioned about the investigation or proof of these comments on the observations.
5. The antennas used are directional and it is observed that when the device and the sensor are not facing each other there is an increase in SNR and the results are degraded. However, this test is done on two corner cases which I believe should also be done on extreme cases like the device facing totally opposite to the sensors, etc. The results could be not promising but in a real-time setting, the audio device can be in. any direction to the sensor.
6. The "Sound Source Frequency" test is done at a very short distance (1m) from the device. I believe the increase in the distance there will be an increase in SNR and the claimed accuracy might not hold true.

7. The test doesn't extend to the recovery of the human voice and is performed only on songs (majorly poems, which are slow but high tunes), and considering UWHear has a low sampling rate, accuracy can be a question.

Other limitations are well mentioned in the paper and some of them are kept for Future Work.