Your Smart Speaker Can “Hear” Your Heartbeat!

by Fusang Zhang, Zhi Wang, Beihong Jin, Jie Xiong, Daquing Zhang

Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (December 2020)
Contactless Vital Sign Monitoring

• Traditionally, monitoring respiration and heart rate requires dedicated equipment and professional training

• **Pros:** Highly accurate

• **Cons:** Requires a trained operator, requires on-site use (or heavy investment), requires physical contact
Ubiquitous Human Sensing

Taxonomy for Human Sensing

Research Question

Can we use a commodity smart speaker to monitor respiration and heartbeat in a contactless, home audio approach?
Major Contributions

1. **First to exploit commodity smart speakers** for accurate heartbeat monitoring*

2. **Demonstrates the system** in real-time with robustness to various real-world concerns.

3. **Proposes a “virtual signal” delay-removal scheme** to deal with random system delay present in smart speakers (relevant to any smart-speaker-based acoustic sensing)

4. Proposes a model to **separate heartbeat and respiration** signals
Let’s Dive In!
State of field: Vital Sign Monitoring

• Standard practice respiration/heartbeat monitoring requires trained operators and special equipment
  • Respiration – impedance photoplethysmography (PPG)
  • Heartbeat – electrocardiogram (ECG)

• At home use: portable devices and wearables
  • Often expensive
  • Often less accurate
State of field: Home Audio Approach

• Success with contactless respiration monitoring
  • ~5mm chest displacement

• **Research gap:** Very little work on heartbeat monitoring
  • 0.1–0.5 mm chest displacement (order-of-magnitude smaller!)
  • Superimposed over the respiration signal

<table>
<thead>
<tr>
<th>Vital signs and body motion</th>
<th>Displacements</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Rate</td>
<td>0.1-0.5mm</td>
<td>0.8-2Hz</td>
</tr>
<tr>
<td>Breathing Rate</td>
<td>1-5mm</td>
<td>0.1-0.5Hz</td>
</tr>
<tr>
<td>Slow body movement (torso)</td>
<td>Decimeter level</td>
<td>0.1-2Hz</td>
</tr>
<tr>
<td>Fast body movement (hand/leg shake)</td>
<td>Centimeter level</td>
<td>0.5-5Hz</td>
</tr>
</tbody>
</table>
State of field: Other highly relevant work

• Acousticcardiogram: monitoring heartbeats using acoustic signals on smart devices, IEEE INFOCOM (2018)
  • Uses a smart phone

• Using smart speakers to contactlessly monitor heart rhythms, Nature Communications Biology (2021)
  • https://www.nature.com/articles/s42003-021-01824-9
  • Extremely similar, focus on medical co-morbidities and BMI
Proposed Solution
Frequency-Modulated Continuous Wave (FMCW) Signals***

Recall Homework 2...

$S_{src} :=$

$S_{recv} :=$

Finding Source Distance:

$$\arg \max S_{src} \ast S_{recv}$$

time delay
distance
Frequency-Modulated Continuous Wave (FMCW) Signals

Key intuition: by sending a signal with a time-varying frequency, it’s easier to pattern match against the received signal.

For formal description, see paper. For tutorial, radartutorial.eu looks pretty good!
Technical Challenges

1. **Random buffer delay.** Smart speakers have a random time delay before signal transmission, causing ordering issues in the received signal.

   ✔ Construct a “virtual signal” at the right start time
Constructing a “Virtual” Signal

1. Mix transmitted and received signals
   Result is one of three cases:
   1. Small system delay between 0 and $T/2$
   2. Large system delay between $T/2$ and $T$
   3. System delay of approximately $T/2$

2. Find the direct signal (strongest peak) with delay $t_w$

3. Construct the virtual transmission signal: $x'(t) = A \cos(\phi(t + t_w))$

4. Set the starting time to 0

For detailed algorithm + equations, see paper.
FMCW Processing Module

1. Transmit FMCW signal from speaker
2. Use direct path + received signal to construct a virtual signal
3. Use virtual + reflected signal to find a chest displacement signal (with superimposed modes)
4. Send on to the next module!
Technical Challenges

1. **Random buffer delay.** Smart speakers have a random time delay before signal transmission, causing ordering issues in the received signal.
   ✓ Construct a “virtual signal”

2. **Heartbeats result in tiny chest motions.** Motion from heartbeats (0.1–0.5 mm) is much smaller than respiration (~5 mm) and can get lost in ambient noise.
   ✓ “Complete ensemble empirical mode decomposition” to separate respiration and heart rate from ambient noise
Superimposed Signals

- **Key intuition:** the displacement (from respiration or heartbeats) will induce a *phase change* in the received signal.

Fig. 9. The resultant signals of superimposed heartbeat and respiration motions.
Heartbeat Extraction: Separation

• We need to decompose these signals!
• **Basic approach:** Empirical mode decomposition (EMD) into intrinsic mode functions (IMF) though a recursive process called “sifting”
  • Find the local max and min of superimposed signal
  • Find the upper and lower envelopes (bounding curves) with cubic splines, then average
  • Check if the IMF condition is satisfied (base condition), else repeat
• **Actual approach:** complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN)
  • Adds an ensemble approach + Gaussian noise
  • Helps reduce mode mixing with non-uniform extremum distribution

For detailed algorithm + equations, see paper.
Heartbeat Extraction: Separation

![Graph showing decomposed phase change signals]

Fig. 10. The decomposed phase change signals.
Putting it all together
Evaluation

- Each test ran 5 minutes, repeated twice

- Baselines:
  - **ACG** – smartphone-based “acousticcardiogram”, 20 cm from subject
  - **Kiwi** – contact-based smartphone app (camera+finger-based)
  - **Kangyuan band** – contact-based smartphone app with band (looks like a Wahoo or Garmin band)

- Metrics:
  - Median HR error
  - Median IBI error
## Basic

<table>
<thead>
<tr>
<th>Method</th>
<th>HR error (bpm)</th>
<th>IBI error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>This Paper</td>
<td>0.75 bpm</td>
<td>1.73%</td>
</tr>
<tr>
<td>Kangyuan band</td>
<td>1.01 bpm</td>
<td>1.57%</td>
</tr>
<tr>
<td>ACG</td>
<td>3.07 bpm</td>
<td>3.1%</td>
</tr>
<tr>
<td>Kiwi</td>
<td>3.24 bpm</td>
<td>3%</td>
</tr>
</tbody>
</table>

Performance is on-par with the chest band approach.
Technical Challenges

1. **Random buffer delay.** Smart speakers have a random time delay before signal transmission, causing ordering issues in the received signal.
   ✓ Construct a “virtual signal”

2. **Heartbeats result in tiny chest motions.** Motion from heartbeats (0.1–0.5 mm) is much smaller than respiration (~5 mm) and can get lost in ambient noise.
   ✓ “Complete ensemble empirical mode decomposition” to separate respiration and heart rate from ambient noise

3. **This system should work in the real world.** Need to account for practical conditions like different locations, poses, ambient noise, and user demographics.
This system should work in the real world?

How might this be more complicated in the real world?

• Different user demographics
• Different sitting/lying postures
• Different speaker-relative positions
• Changing heart rates
• Looser clothing
• Ambient noise
What about age/gender diversity?
What about different postures?

(a) Sitting scenario

(b) Lying scenario
What if the heart rate isn’t constant?
What about varied clothing?

Fig. 20. Different clothing scenarios.

Fig. 21. Impact of different clothing.
What about different distances?
What about different directions?

Fig. 23. Impact of user-device sitting direction.

Fig. 24. Impact of user-device lying direction.
What about realistic environment noise?

**Note:** ultrasonic signal (16–21 kHz), so we can filter out “audible” noises (e.g. talking, music, etc.) with a highpass filter.
Reflections
Is this practical?

• Kind of! Requires explicit, intentional use
• Several major limitations:
  • Doesn’t work for typing, talking, movement
  • Need to be alone in the room
• Other limitations? (time-permitting)
Security and Privacy

- Highpass filter for audible frequency bands
- Implicit consent from use requirements (staying still by yourself in short range)
- Data sharing and alternate uses
- Device security
- Any thoughts? (time-permitting)
Take-Aways

- Demonstrates accurate, real-time acoustic heartbeat monitoring using commodity hardware
  - Robust to many real-world conditions
  - Significant practical limitations (requires intentional use)
- Uses signal processing techniques useful beyond this paper
  - Dealt with practical smart speaker behavior inconsistencies
  - Designed FMCW signals with linear chirps
  - Mode decomposition to separate signals
- Intended as part of a larger ubiquitous health sensing ecosystem (vital sign monitoring, fall detection, gesture recognition, etc.)
Your Smart Speaker Can “Hear” Your Heartbeat! (2020)

https://dl.acm.org/doi/pdf/10.1145/3432237