ARSteth: Enabling Home Self-Screening with AR-Assisted Intelligent Stethoscopes

Kaiyuan Hou
Columbia University
New York, New York, USA
kh3119@columbia.edu

Stephen Xia
Columbia University
New York, New York, USA
sx2194@columbia.edu

Emily Bejerano
Columbia University
New York, New York, USA
eg3205@columbia.edu

Junyi Wu
Columbia University
New York, New York, USA
jw4173@columbia.edu

Xiaofan Jiang
Columbia University
New York, New York, USA
jiang@ee.columbia.edu

ABSTRACT

The stethoscope is one of the most important diagnostic tools used by healthcare professionals, through a process called auscultation, to screen patients for abnormalities of the heart and lungs. While there are digital stethoscopes on the market which ease this process, it still takes years of training to properly use these devices to listen for abnormal sounds within the body. We present ARSteth, an intelligent stethoscope platform that improves the accessibility of stethoscopes for the general population, allowing anyone to perform auscultation in the comfort of their own homes. Our platform utilizes a combination of augmented reality (AR), acoustic intelligence, and human-machine interaction to dynamically guide users on where to place the stethoscope on different parts of the body (auscultation points), through visual and audio cues. Through user studies, we show that ARSteth, on average, can guide users within *13.2* mm from optimal auscultation points marked by licensed physicians in *13.09* seconds for each auscultation point. By guiding users towards more effective auscultation points, make preventative health screening more accessible and effective for everyone we are able to achieve higher confidence on classifying heart murmurs.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; • **Computer systems organization** → *Embedded and cyber-physical systems.*

KEYWORDS

digital stethoscope, human computer interaction, smart home

ACM Reference Format:
Kaiyuan Hou, Stephen Xia, Emily Bejerano, Junyi Wu, and Xiaofan Jiang. 2023. ARSteth: Enabling Home Self-Screening with AR-Assisted Intelligent Stethoscopes. In *The 22nd International Conference on Information Processing in Sensor Networks (IPSN '23), May 09–12, 2023, San Antonio, TX, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3583120.3586962

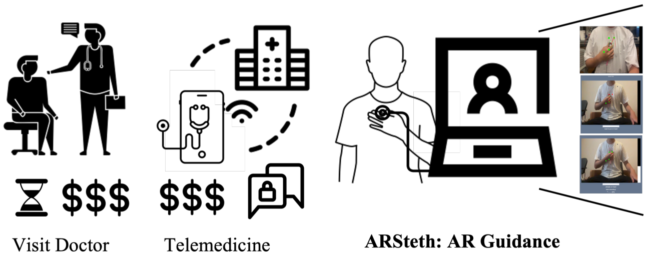


Figure 1: Currently, screening for heart related diseases require physical doctor visits, which is costly. Digital stethoscopes are widely accessible, but still often cost hundreds of dollars and require years of training to use properly. ARSteth enables users to self-screen at home by following intuitive AR-assisted guidance.

1 INTRODUCTION

Inaccessibility to health care, more specifically preventative services, is a major problem in the United States and around the world. It is recommended to have a comprehensive physical examination at least once per year [15]. However, preventative care is often prohibitively expensive, with more than 26,000 Americans dying each year due to lack of health insurance [26]. This problem is exacerbated in developing countries where more people are unable to pay the high cost of medical expenses, leading to worse outcomes as diseases are often not detected until later stages [1, 25]. Many systems have been developed to improve and enable user safety [33, 34], privacy [32], and preventative health screenings from both the mental [19, 21–23] and physical [12, 13] perspectives to improve quality of life and reduce the strain on healthcare infrastructure. There are also many commercial devices that provide quick medical insights include mercury and infrared thermometers, blood pressure monitors, and glucometers. However, household devices to monitor one's heart condition are lacking. With cardiac disease being a leading cause of death worldwide, a widely accessible solution is needed to allow everyone to monitor heart health without requiring access to a trained medical professional [35].

IPSN '23, May 09–12, 2023, San Antonio, TX, USA

ACM ISBN 979-8-4007-0118-4/23/05…$15.00
https://doi.org/10.1145/3583120.3586962

The stethoscope is the hallmark non-invasive tool used by healthcare providers to examine a person's circulatory and respiratory systems by listening to the sounds produced internally by the body in a process called auscultation. During a typical examination, the physician places the diaphragm of the stethoscope at various locations on the patient's body (typically areas on the chest, abdomen, and back) and listens for sounds that indicate abnormal functioning, such as heart murmurs, fluid in lungs, or abnormal bowel movements. The results from this examination, along with the patient's medical history and other provided information, are used by the healthcare provider to determine if further examination and/or treatment is necessary.

There are several professional-grade digital stethoscopes on the market which can interface with smartphones and laptops, allowing anyone to perform auscultation. Even though digital stethoscopes are widely accessible, an average non-licensed person would not be able to screen themselves because it takes years of professional training to learn what sounds are ab/normal and where to place and adjust the stethoscope.

We propose and design a non-invasive digital stethoscope platform, ARSteth, powered by artificial intelligence (AI) with the goal of enabling the general population to screen themselves for abnormal heart. ARSteth utilizes computer vision and acoustic algorithms to identify effective auscultation points where machine learning detectors can better distinguish between normal and abnormal heart sounds. Additionally, ARSteth guides users on where to place the stethoscope, through visual and audio cues, using live video from the user's computer camera and the sounds recorded by the stethoscope. ARSteth represents a step towards enabling universal health screening, allowing underserved and low-resource communities access to a powerful health screening tool without requiring access to a medical professional.

In this paper, we make the following contributions:

- We propose ARSteth, an AI-based stethoscope platform that improves accessibility to low-cost health screening. ARSteth guides users in placing the stethoscope at different auscultation points through real-time audio-visual feedback. ARSteth analyzes sounds recorded from the stethoscope at each auscultation point using signal processing and machine learning techniques to screen for heart murmurs.
- We introduce a method to locate auscultation points by combining the advantages of both video and audio to overcome the constraints of either sensing modality alone. Mimicking a physician, ARSteth leverages sight (computer vision) to estimate the coarse location of one's heart and then fine-tunes the location with the sounds recorded by the stethoscope in real-time. In doing so, ARSteth compensates for variations due to different body shapes from different users.
- We suggest modeling the auscultation points from a qualitative description to a quantitative representation. We determine the location of auscultation points with respect to the coordinates of the shoulder positions of the user. This allows us to find initial auscultation points quickly.
- Through user studies, we demonstrate that ARSteth, on average, can guide users within *13.2* mm from optimal auscultation points marked by licensed physicians in *13.09* seconds, without the presence of a medical professional.
- We show that ARSteth greatly improves the usability compared to existing state-of-art stethoscopes, scoring 12.6% higher rated on the Likert scale across on four aspects of usability through a usability study.

2 RELATED WORKS

Stethoscopes are perhaps one of the most well-known health screening tools used by primary care physicians to listen for abnormal sounds, indicative of illness, coming from the within the body. Recently, there has been a growth of digital stethoscopes on the market, as part of an effort to provide both doctors and the general population easy access to a powerful health screening tool.

There are several stethoscopes on the market that provide automatic heart murmur detection and amplified auscultation sounds with artificial intelligence, including the 3M Littman Core Digital Stethoscope [7]. However, our own experiments with this stethoscope yielded a high rate of false positive murmur detections. There are also several stethoscopes and that help guide users on where to properly place a stethoscope [3]. However, none of these works dynamically guide users in real-time and only display an image showing where to place the stethoscope on a static model.

An alternative method for monitoring the heart is to use devices other than stethoscopes, which can assist in virtual visits (telehealth) [14, 15]. However, this approach still presents a challenge regarding the placement of the stethoscope. [4, 11]. However, this approach still presents a challenge regarding the placement of the stethoscope. Properly using a stethoscope requires years of training to understand where to place the stethoscope and which types of sounds to listen for.

The goal of our work is to improve the usefulness of stethoscopes for the general population besides virtual visits (telehealth). By incorporating dynamic AR guidance, an untrained person can easily and quickly screen his/her health with limited assistance from a healthcare provider. The benefits of incorporating AR and vision-based guidance has been shown to improve accessibility in other areas of medicine [2].

Many works haven been done to classify various sounds recorded from a stethoscope into different types of heart murmurs and lung ailments [5, 9, 10, 16, 27–29].In this work, we show that by incorporating dynamic AI-driven AR guidance that aids users in where to properly place the stethoscope during the examination, we can more quickly obtain higher quality heart that improve the performance of algorithms for detecting heart murmurs ailments.

3 AUSCULTATION POINTS MODELING

The first task to address is to identify the auscultation points on the body. As shown in Figure 2, physicians routinely auscultate four points to listen to four heart valves: Aortic, Pulmonary, Tricuspid, Mitral [31]. The identification of auscultation points is typically dependent on the subject's individual body structure. During auscultation, physicians first assess the subject's body to determine the

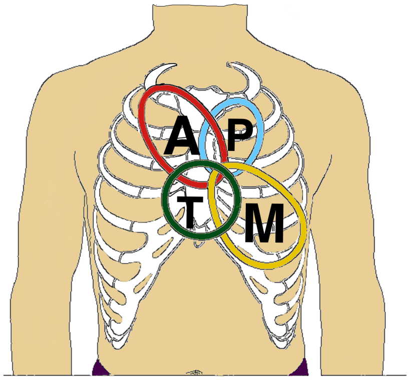


Figure 2: Typical auscultation points. *Aortic (A)*: right side of the sternum; *Pulmonary (P)*: left-hand side of the sternum; *Tricuspid (T)*: in the fourth intercostal space, along the lower-left border of the sternum; *Mitral (M)*: in the fifth intercostal space, along the mid-clavicular line[14].

location of the sternum and intercostal muscles, which are associated with the auscultation points. They then palpate the chest area around these locations to locate the auscultation points and place the stethoscope accordingly. However, due to the presence of bones and unique body structures, standard auscultation locations may not always yield the best quality sounds. Therefore, physicians often make slight adjustments to the stethoscope position to enhance the quality of the sounds.

The standard procedure for auscultation involves physicians relying on information from visual, tactile, and auditory senses to locate the specific points on the body where they need to place the stethoscope to hear relevant sounds. ARSteth leverages multiple sensing modalities to replicate this procedure while guiding users. However, identifying the precise location of auscultation points is challenging due to their variability among individuals, making it heavily reliant on the physician's expertise. Thus, one of the most challenging aspects of self-screening with a stethoscope is the difficulty associated with finding the auscultation points. A purely computer-vision-based approach is insufficient because clothing commonly obstructs details of the body.

3.1 Determining Coarse Auscultation Locations

In most auscultation tutorials and guides, the location of auscultation points is typically shown on a skeleton diagram, with a spatial relationship between the points and the skeleton indicated. Apart from the sternum and intercostal spaces, the positions of the left and right shoulders can also be used as reference points. We can adopt computer vision-based method to quantitatively determine the locations of auscultation points on a person's body by establishing a coordinate system with the left and right shoulder positions as references.

Using the shoulders to locate the auscultation points has several advantages. First, there has been significant research on locating the shoulders in the context of augmented reality, which provides a well-established foundation for our approach. Second, since the shoulders represent the outline of the body, movements of the upper body forward or backward do not affect the detection of the auscultation points when we anchor them to this plane. On the other hand, if we were to use a landmark that is not on the body's outline as a reference for locating the auscultation points, the coordinate system would be distorted by variations in the person's sitting position, leading to errors in the estimation of the auscultation points. Finally, the shoulder information can be used to determine the subject's sitting position, which can aid in the auscultation process.

Assume that the right shoulder is located at (x_r, y_r, z_r) and the left shoulder is located at (x_l, y_l, z_l) in 3D space. We create a coordinate system where the origin is the right shoulder, the x-axis points from the right shoulder to the left shoulder, and the y-axis points downwards and is perpendicular to the x-axis. We represent the locations of the auscultation points as follows:

$$\begin{aligned} x_i &= x_r + (C_{ix}/C) * (x_r - x_l) \\ y_i &= (y_r + y_l)/2 + (C_{iy}/C) * (x_r - x_l) * Y/X \end{aligned} \tag{1}$$

Here, (x_i,y_i) represents the coordinates of the ith auscultation point. C_{ix} and C_{iy} are parameters for each auscultation point that will be explained in Section 3.1.1. C is a constant with an empirical value of 1210, which represents the length of the shoulders. X and Y are the aspect ratio of the image.

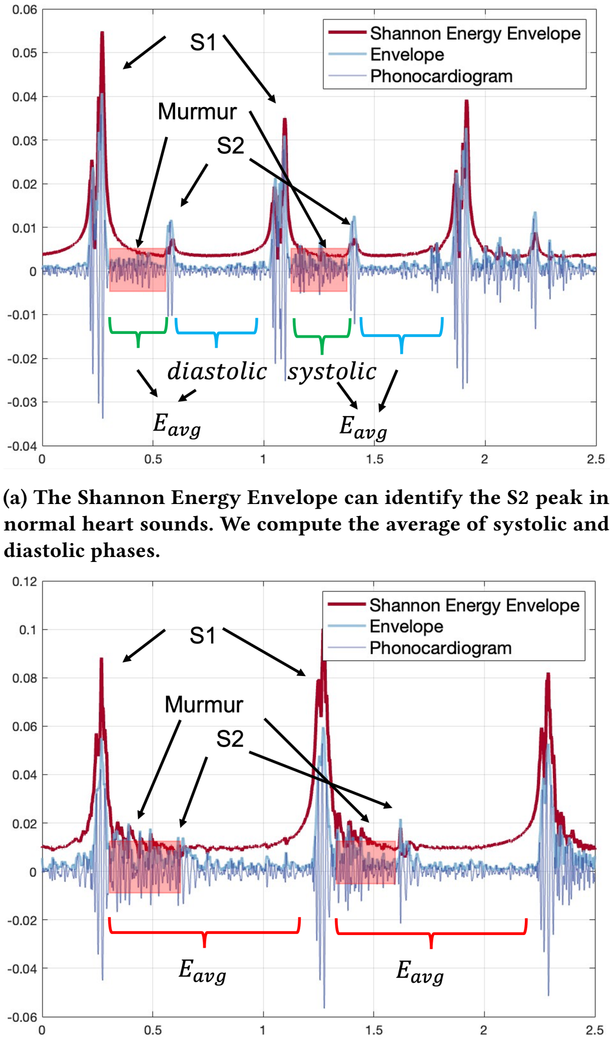
3.1.1 *Auscultation Point Parameters:*

To estimate auscultation parameters in Equation 1 (C_{iy}, C_{iy}), we recruited a group of individuals. A physician marked the locations of four auscultation points on each subject using stickers. These stickers were then detected by a program, and the center of the resulting bounding boxes was computed. Using Equation 1, we solved for parameters C_{iy} and C_{iy} across all auscultation points. For each individual, there were four tuples representing the x and y parameters for the four auscultation points.

Next, all combinations of these tuples from all individuals were traversed, with each combination forming a new set of four tuples. These sets of parameters were used to place stickers on a subject, and the distance error between the auscultation points found using the formula and the auscultation points marked by the physician was measured. We integrated the set of parameters that gave the minimum error across all subjects into ARSteth. This final parameter resulted in an average distance error of *2.88* cm across all subjects.

3.2 Fine-Tuning Auscultation Points

After the coarse location of auscultation points are obtained, we discuss how to update the auscultation point parameters by analyzing the sound captured from the stethoscope. A logical question that arises is: why are vision estimates not enough? First, we determine the points of auscultation by referencing the position of two shoulders, which has been discussed in the previous section. Nevertheless, this introduces two potential sources of error: there is a certain level of uncertainty with respect to the estimated landmarks, and the equation used to derive auscultation points relies on parameters that do not account for differences in body shape; consequently, the estimated points obtained solely by computer



(a) The Shannon Energy Envelope can identify the S2 peak in normal heart sounds. We compute the average of systolic and diastolic phases.

(b) S2 cannot be clearly found in severe heart murmurs.We compute the average between two consecutive S1 sounds.

Figure 3: Shannon Energy Envelope, Envelope and Original Phonocardiogram(PCG) with mild and severe murmur.

vision are frequently off by a few centimeters. Second, the placement of the stethoscope affects the quality of auscultation. Better estimated auscultation points often significantly increase the characteristics of abnormal sounds. Due to **radiation**[1], the auscultation point for the heart is more crucial to detect heart murmur. We may detect murmur at multiple locations, but we are mostly interested in the location with the highest intensity; it is often difficult to detect abnormal sounds at locations far from optimal auscultation points. This is due to their low signal-noise ratio (SNR) or due to attenuation during propagation to the stethoscope.

A murmur will introduce extra peaks or non-zero peaks in the phonocardiogram (PCG) during the systolic phase or diastolic phase. It will also cause an extra lobe in the Shannon Energy Envelogram[17]. We compute the Shannon Energy E as:

$$E = -x^2 \cdot log(x^2) \tag{2}$$

Sound can be segmented using a Shannon Energy envelogram, which accentuates medium-intensity signals and greatly reduces the impact of low-intensity noise and artifacts. We then compute the envelopes' average Shannon Energy during the systolic phase and diastolic phase, where the envelopes average Shannon Energy E_{avg} can be computed as:

$$E_{avg} = -\frac{1}{N}\sum_{i=1}^{N} x_i^2 \cdot log(x_i^2) \tag{3}$$

where N is the number of samples. The systolic and diastolic phases can be easily found by locating the end and start of S1 and S2 sounds which are always the highest peaks in envelopes as shown in Figure 3a. In this work, we are more interested in the start and end times of the S1 and S2 sounds instead of the segments. S1 and S2 can be found by setting thresholds. However in some cases with severe heart murmur, the intensity of murmur sound can be greater than S2 sounds such as in Figure 3b. We then only focus on the occurrence of S1 sounds and compute the average between two consecutive S1 sounds.

Algorithm 1 Fine-Tuning of Target Auscultation Point

```
Input: loc: array of locations $(x, y)$, energy: array of corresponding
    computed average energies
Output: position: tuple representing final location of target auscultation point
while True do
    if energy is ascending then
        if arrive at target auscultation point then
            Append a counterclockwise trajectory to path
        else
            Append current location to loc
        end if
    else
        position $\leftarrow$ loc[-1]
        break
    end if
end while
Compute the percentage change between the final $(x, y)$ and
    the initial $(x, y)$ at this auscultation point
for each unexamined auscultation point do
    Scale the initial $(x, y)$ of the unexamined point with the
        percentage change to reflect the variance in body shapes
end for
```

Algorithm 1 details ARSteth's auscultation point fine-tuning process. This algorithm computes the average Shannon Energy when the stethoscope is in close proximity to the target auscultation point and leverages this to fine-tune the location of the target auscultation point in the same manner of a physician. This algorithm takes an array of locations loc and an array of corresponding

[1] Murmurs are typically most prominent at a particular listening post. They frequently radiate to other listening posts or body parts. For instance, aortic stenosis commonly radiates to the carotid arteries, whereas mitral regurgitation radiates to the left auxiliary region.

computed average energies energy as input. It fine-tunes the target auscultation point by repeating the process of sampling and computing the recorded sample until a decrease in energy is observed. It then reassigns the auscultation point to the previous location that gives a higher value in terms of energy. If the energy is always increasing and the user has positioned the stethoscope, the algorithm extends the path with a counterclockwise trajectory until a local maximum is found. This procedure ensures that we can find the best auscultation points with limited known knowledge. Once the final location of the target auscultation point is determined, the algorithm computes the percentage change between the final and initial (x, y) at this auscultation point. The initial (x, y) of the unexamined auscultation points are then scaled with this percentage change to reflect the variance in body shapes.

4 IMPLEMENTATION

Figure 4 is ARSteth's system architecture. ARSteth takes the camera, stethoscope, and several user inputs. In this work, we employ Mediapipe[20] to determine the landmarks of body such as the shoulder coordinates in the image and the stethoscope location held by thumb and index. The user inputs height, weight and gender, which is used to generate the initial auscultation point parameters. Sound observed by the stethoscope is used for updating the parameters or terminating the guidance to start recording. ARSteth generates a detailed report that shows what and where the abnormalities are detected. We first illustrate how ARSteth assists in the detection of auscultation points. We also examine the interaction between ARSteth and humans. Furthermore, we introduce how ARSteth calibrates auscultation points according to sitting position, the disease classification method, and our custom-designed stethoscope.

4.1 Guidance and Human Interaction

The identification of auscultation points requires human-computer interaction which prompts the user to adjust the stethoscope as the estimated auscultation points are dynamically updated according to Algorithm 1.

4.1.1 *Sitting Position Regulation.*

During a physical examination, patients are not required to assume a particular position. However, during self-screening, the predicted points of auscultation can vary depending on the sitting position, as shown in Fig. 5. The green dots on the figure represent the estimated points of auscultation, and different sitting positions can significantly affect the accuracy of the auscultation points. Our preliminary experiments show that maintaining a stable position and gesture during the test can improve sound quality. Therefore, it is crucial to ensure that the user is sitting properly.

Two factors need to be considered to achieve this. First, we need to determine whether the user is sitting upright, and second, we need to determine whether the user is facing the camera. Since we intend to use augmented reality to assist the user's self-screening, the user's upper body is visible in the camera's field of view, providing the necessary information to verify the user's sitting posture. Suppose the position of the right shoulder is (x_r, y_r, z_r), and the position of the left shoulder is (x_l, y_l, z_l).

To determine if the user is sitting upright, we calculate the angle displacement from the roll axis of the user with respect to the horizontal line using $\arctan \frac{y_l - y_r}{x_l - x_r}$. If the resulting angle is less than a specified threshold, we prompt the user to lower their left shoulder, as shown in Fig. 5a. Conversely, if the angle exceeds a certain degree, we ask the user to lower their right shoulder.

To determine if the user is facing the camera, we find the angle displacement from the yaw axis of the user with respect to the normal direction of the camera. We calculate the difference in the heights of the shoulders to estimate this displacement, i.e., $z_l - z_r$. If the difference is less than a specified value, we ask the user to turn left, as shown in Fig. 5b. If the difference exceeds a certain value, we ask the user to turn right. The threshold for both cases will be presented in Section 5.2.

4.1.2 *Stethoscope Positioning.*

In some circumstances, individuals may cover the drum of the stethoscope with their hand, resulting in the loss of tracking when a model of the stethoscope directly trained for detection is used. To mitigate this, we track the thumb and index finger and approximate the stethoscope as a circle with a diameter corresponding to the distance between these two coordinates.

4.1.3 *Logic of Guidance.*

ARSteth provides two types of guidance. The first type is the sitting position regulation, as discussed in Section 4.1.1. The second type directs the user to move the stethoscope to find the auscultation points, similar to how a physician would. To guide the user in how to sit and adjust his/her body, the video stream from the camera is mirrored. Once the user meets the requirements, an image of the heart and four coarse auscultation points appear on the body, which is comparable to a physician's initial estimation. The user is then guided to place the stethoscope at the four auscultation points (in order): *Aortic*, *Pulmonary*, *Triuspid*, and *Mitral*.

During the process, ARSteth detects the stethoscope and shows a flashing red arrow from the stethoscope to the current targeted auscultation point. ARSteth also samples the sound from the stethoscope and performs pre-processing on the acquired sounds. The frequency spectrum of heart sounds is typically between 10 and 200 hertz. S1 has a lower pitch and is a longer-lasting sound, while S2 is a shorter-lasting sound with a higher pitch. S1 dominates the 10-140 hz range, whereas S2 dominates the 10-200 hz band. S3 and S4 have low amplitude and frequency, between 20 and 70 hertz. If there is a dysfunction in the cardiac system, such as murmurs or mitral valve stenosis, the spectral content may increase to 600-700 hz. To preserve information on heart features, ARSteth low-pass filters the signal with a cut-off frequency of 1 kHz to remove environmental noise and possible noise between the stethoscope, skin, or clothes.

To monitor the movement of the stethoscope, ARSteth saves the location of the center of the stethoscope in a queue for updating the location of the auscultation points when Algorithm 1 terminates. Once the algorithm finishes, the locations of the remaining unexamined auscultation points are updated. Ideally, ARSteth wants the user to move the stethoscope slowly. If the movement is too fast, ARSteth will ask the user to move the stethoscope in a circular path around the original location to obtain additional samples. This is a

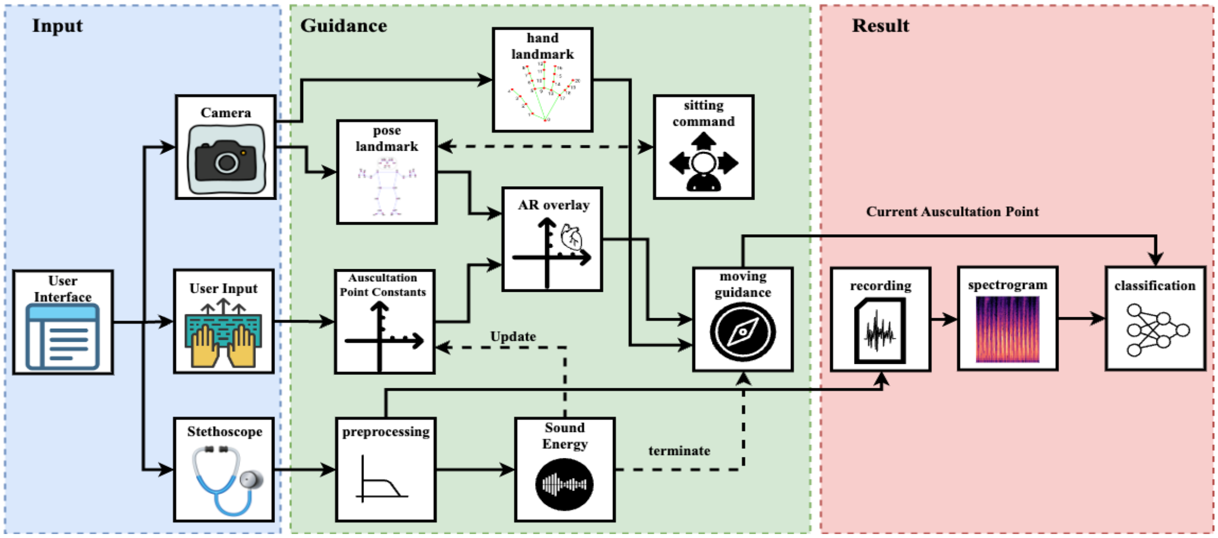
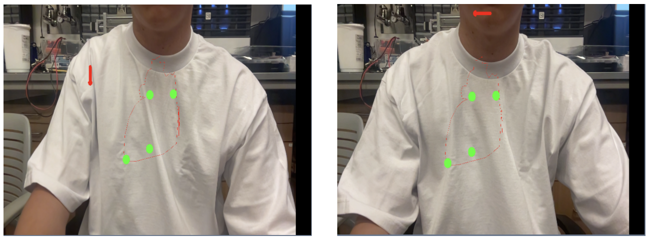


Figure 4: System Architecture block diagram. Three inputs (camera, user input and sound from the stethoscope) are used to guide the user and assess the potential heart ailments.



(a) Left shoulder is too high **(b) Turn left**

Figure 5: Incorrect sitting position causes difference in coarse estimation. Left: the left shoulder is too high. Right: the right shoulder is too close to the camera.

dynamic procedure where the user follows the arrow to the next targeted position displayed, and ARSteth updates this position until the criteria are met.

The center of the stethoscope is transformed into the same coordinate system as auscultation point. We compute the Euclidean distance between the stethoscope and the fine-tuned auscultation points, if the distance is less than 30 pixels, ARSteth will start recording the sound for 5 seconds. To guarantee the quality of the sample, we ask the user to maintain the stethoscope's position. During the acquisition phase, if the distance between the stethoscope and the target position exceeds the threshold, the sound collection will terminate, and guidance will repeat.

4.2 Auscultation Point Calibration

The user's sitting position can greatly affect the accuracy of the auscultation points. To reduce errors in auscultation point locations caused by different positions, ARSteth guides users to sit straight and upright, as discussed in Section 4.1.1. To further reduce errors and variations caused by the plane of the body not being completely parallel to the plane of the camera lens, we perform a projection discussed below.

We model the auscultation points with respect to left and right shoulders by Equation 1. The parameters for each point of auscultation of the heart are based on the real body, which may be rotated with respect to the camera, but not the flattened 2D body in the image. The variance in coarse location is coming from the lack of depth in the 2D graph. In the case when the user's body is tilted with their right shoulder closer to the camera, one line segment closer to the right shoulder will be shorter than a line segment with the same length in image. This is why we need to perform projection to the coordinate system whose x-axis is parallel to the horizontal line as shown in Figure 6. We want to find the actual length in the xyz coordinate system given the information in $x'y'z'$ coordinate system. We modify the auscultation equation along the x-axis to make it correlated with more information about the body. There are two angles we need to consider, the angle between the shoulder level and the horizontal line θ and the inclination angle that is normal to the plane of the image ϕ, they are calculated by:

$$\begin{aligned} \theta &= \arctan(\frac{y_l - y_r}{x_l - x_r}) \\ \phi &= \arctan(\frac{z_l - z_r}{\sqrt{(x_l - x_r)^2 + (y_l - y_r)^2}}) \end{aligned} \tag{4}$$

The length of the original vector in 3D space represents the visual shoulder length in the 2D flatted plane which is $x_r - x_l$. However, the actual shoulder length is greater than this visual length due to the shoulder not being parallel to the x-axis. The auscultation point parameter on the plane when the body is parallel to the camera. The shrink factor k is calculated as $k = cos(\phi) * cos(\theta)$. We assume that the shoulder that is closer to the camera can allow for more reliable measurements. If the right shoulder is closer, then we still use the location of right shoulder to be the origin of the coordinate

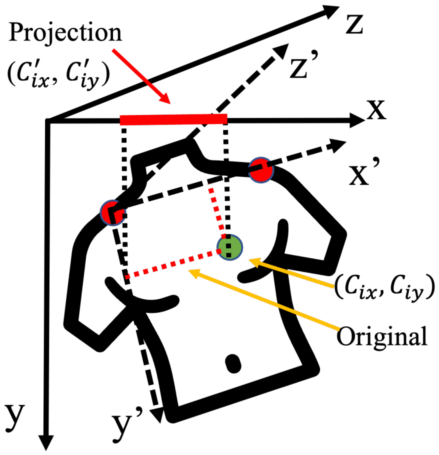


Figure 6: Auscultation point projection. The auscultation point parameters are based on the ratio of shoulder length as indicated with dashed lines. The new parameters are the fraction that projected onto the x-axis which is parallel to the horizontal line as indicated in the red line.

system. The new auscultation points will be:

$$x = x_r + C_x/C * k * (x_r - x_l) \tag{5}$$

If the left shoulder is closer to the camera, with transformation, we can have the calibrated parameter of x be:

$$x = x_r + (C_x/C * k + 1 - k) * (x_r - x_l) \tag{6}$$

4.3 Disease Classification

The final step of ARSteth involves detecting abnormal heart sounds. Due to our lack of access to actual patients, we utilized a heart sound dataset made available by the PhysioNet/CinC challenge 2016 [18]. This dataset comprises 3126 heart sound recordings that were captured in diverse clinical and non-clinical settings, and features both healthy subjects and those with various heart conditions. The heart sound recordings were gathered from the standard four auscultation points: aortic area, pulmonic area, tricuspid area, and mitral area, and each recording was categorized as normal or abnormal.

4.3.1 *Preprocessing.*

We first downsample all the heart sound recordings to 1 kHz and apply a Butterworth band-pass-filter with cutoff frequencies of 25 Hz and 800 Hz to remove the undesired low-frequency artifacts and high frequency noise such as background noises. The filtered heart sounds are then standardized by subtracting the mean and dividing by the standard deviation. For each heart sound recording, we segment the recording into short intervals of single heart beat cycles by the provided annotations in the dataset. Finally, we use zero-padding in the segments that have less than 1000 samples and discard the segments with more than 1000 samples and use the resulting dataset for training and testing our detector (Section 4.3.2).

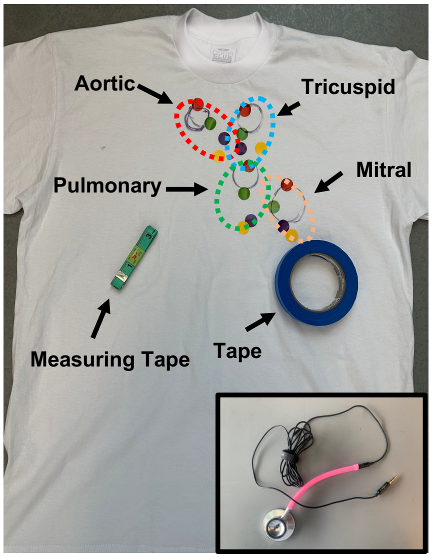


Figure 7: White shirt and customized stethoscope for performing evaluation. In each auscultation point, red dot represents the ground truth, yellow dot represents the initial guess (baseline), purple dot represents the ARSteth without projection, green dot represents ARSteth with projection. Measuring tape is used for measuring the distances on the shirt, and the tape is used to remove the slacks of shirt and reduce the movements of the shirt during the test.

4.3.2 *Neural Network Classifier.*

Many studies have reported the use of neural networks for heart disease classification [5, 9, 10, 16, 27–29]. Since our focus is studying how user guidance can improve ailment detection, we adopt a standard detection model. Our model uses a two-dimensional convolutional (2D-CNN) neural network that takes Mel-frequency cepstral coefficients (MFCC) of each segment after preprocessing as input, and produces a prediction between two classes: normal or murmur. The model achieves an accuracy of 87.18%, a sensitivity of 86.08%, and a specificity of 91.55% on our dataset (Section 4.3.1).

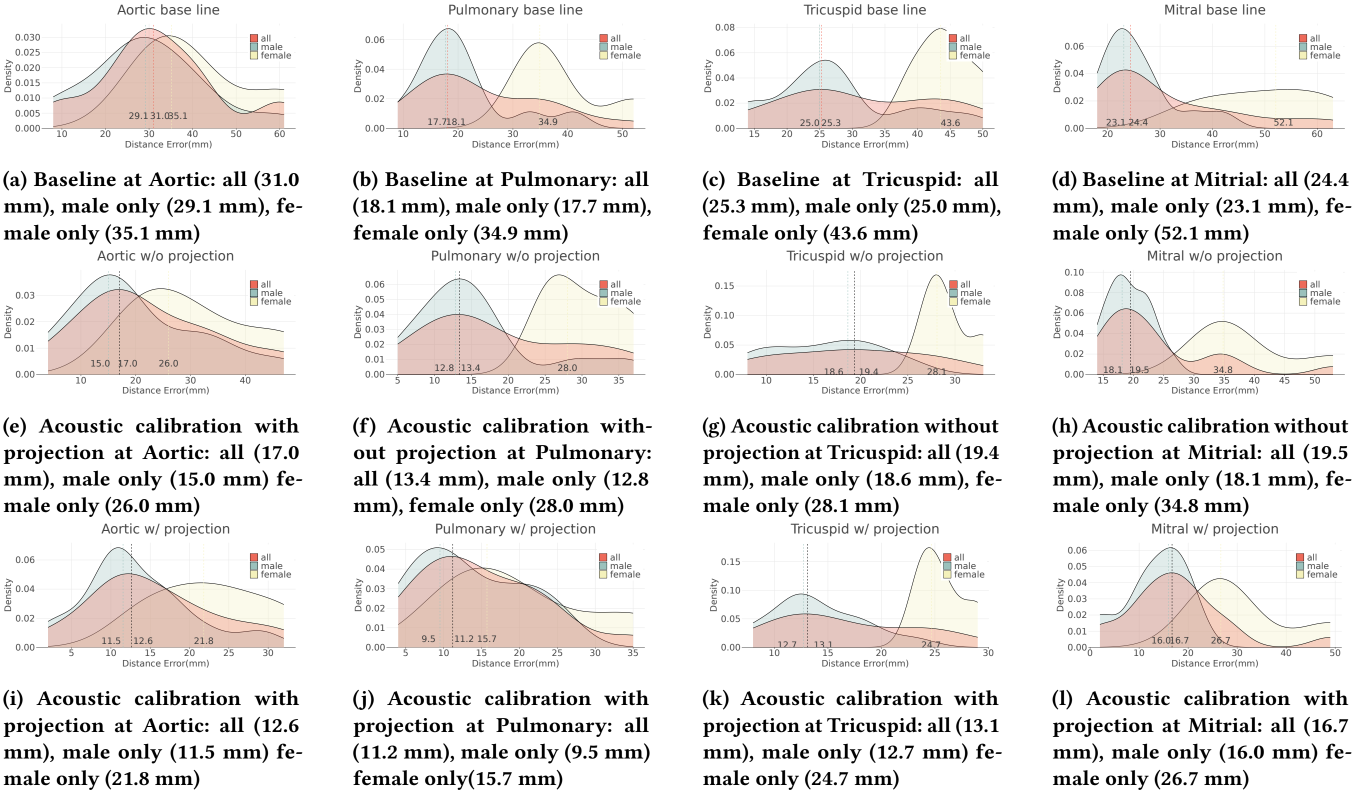
4.4 Customized Stethoscope

We created a custom designed stethoscope with a \$7 dual-head stethoscope and a \$11 microphone from Amazon as shown in the bottom-right corner in Figure 7. We cut the tubing of stethoscope and remove the eartube and eartip, but retain the diaphragm and tubing. We then inserted the microphone partially in the tubing with diaphragm from the drum[2].

5 SYSTEM EVALUATION

In this section, we evaluate the performance of ARSteth from three perspectives. First, we assess the effectiveness of ARSteth in guiding non-professional users through auscultation. Since ARSteth is intended for use at home, its ability to guide users is critical to its

[2] Completely inserting the microphone in the tubing would result in the microphone receiving no sound. We observed no correlation between sound quality and insertion depth after testing various insertion depths.



(a) Baseline at Aortic: all (31.0 mm), male only (29.1 mm), female only (35.1 mm)

(b) Baseline at Pulmonary: all (18.1 mm), male only (17.7 mm), female only (34.9 mm)

(c) Baseline at Tricuspid: all (25.3 mm), male only (25.0 mm), female only (43.6 mm)

(d) Baseline at Mitrial: all (24.4 mm), male only (23.1 mm), female only (52.1 mm)

(e) Acoustic calibration with projection at Aortic: all (17.0 mm), male only (15.0 mm) female only (26.0 mm)

(f) Acoustic calibration without projection at Pulmonary: all (13.4 mm), male only (12.8 mm), female only (28.0 mm)

(g) Acoustic calibration without projection at Tricuspid: all (19.4 mm), male only (18.6 mm), female only (28.1 mm)

(h) Acoustic calibration without projection at Mitrial: all (19.5 mm), male only (18.1 mm), female only (34.8 mm)

(i) Acoustic calibration with projection at Aortic: all (12.6 mm), male only (11.5 mm) female only (21.8 mm)

(j) Acoustic calibration with projection at Pulmonary: all (11.2 mm), male only (9.5 mm) female only(15.7 mm)

(k) Acoustic calibration with projection at Tricuspid: all (13.1 mm), male only (12.7 mm) female only (24.7 mm)

(l) Acoustic calibration with projection at Mitrial: all (16.7 mm), male only (16.0 mm) female only (26.7 mm)

Figure 8: Distance (error) to the actual auscultation points provided by physician. Baseline: coarse estimation only. Acoustic calibration but no projection: coarse estimation + fine tune. Acoustic calibration with projection: coarse estimation + fine tune + projection.

success. Second, we examine the margin of error in the guidance to determine the extent to which ARSteth can tolerate variations in sitting position. This enables us to determine the extent to which ARSteth can guide users accurately and precisely in typical situations. Finally, we present an end-to-end evaluation of ARSteth using simulated abnormalities on a manikin. Prior to conducting the experiments, we obtained Institutional Review Board (IRB) approval and informed consent from 35 participants between the ages of 18 and 44, representing 5 different races. We took measures to protect the privacy of the participants and erased all user activity data after the completion of the study.

5.1 Performance of Auscultation Guidance

The primary goal of this study is to develop an augmented reality (AR) platform that dynamically guides users to the most effective auscultation points. Thus, the system's ability to provide accurate guidance is of paramount importance. This section evaluates the system's performance in terms of timing and accuracy regarding auscultation guidance.

5.1.1 Experiment Setup.

In this section, we describe the setup for our experiment to evaluate the precision of auscultation guidance provided by ARSteth in a quiet room with an environment noise level of 43.6 dB in order to simulate the intended usage of ARSteth in a home environment. To determine the accuracy of the system, we measure the geometric distance between the actual auscultation points and the locations identified by ARSteth. We asked a licensed physician to mark the ground truth locations for all subjects using a standard auscultation procedure as described in Section 3. We provided a shirt and the physician marked the **ground truth** locations with a **red** sticker, as shown in Figure 7.

Each subject used our platform twice while **sitting upright and with minimal shoulder rotation**, with and without projection calibration, as described in Section 4.2. First, we guided the subjects using ARSteth without projection calibration, then marked the identified locations with a **purple** sticker, and marked the initial locations found by computer vision only with a **yellow** sticker as the **baseline**. Then, we enabled projection calibration and repeated the test, marking the new locations with a **green** sticker.

To minimize measurement error caused by shifts in clothing, we tightly secured the shirt around the waist using tape. We measured the distances between the green and red stickers, and between the purple and red stickers, for each cluster representing an auscultation point of the heart. These distances were measured using a tape measure after the subject's shirt was removed and the surface was flattened.

In addition to placing stickers, we recorded the time it took for each subject to find the auscultation points. The timer was started when the subject picked up the stethoscope and began moving and was stopped when the acoustic calibration was terminated.

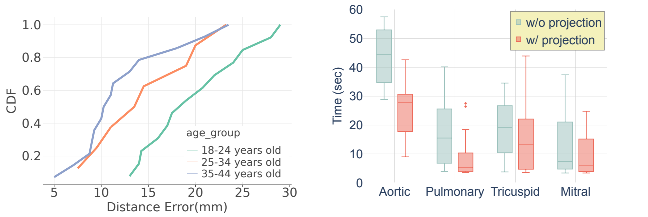
5.1.2 *Precision of Auscultation.*

In Figure 8, it is clear that the baseline method performed the worst, while the acoustic calibration with projection method performed the best. The distance errors for all four auscultation points are consistently lower for both male and female patients when using the acoustic calibration with projection method. The average distance error for all gender is reduced from 24.7 mm in the baseline method to 13.2 mm in the acoustic calibration with projection method, a significant improvement of about 50% (46.57%).

We can also observe that the distance error is consistently smaller for the pulmonary auscultation point compared to the other three points in all methods and for both genders. This could be because the pulmonary valve is closer to the skin surface, making it easier to detect using a stethoscope. We also found that the performance on male patients is generally better than on female patients, which could be due to anatomical differences in the chest wall or other physiological factors affecting the propagation and detection of heart sounds.

We consulted with licensed physicians to validate the accuracy of the auscultation points identified by our system. The physicians found the sounds collected at those points to be acceptable and within the range of corresponding main heart valve areas. However, it is important to note that the diameter of an auscultation point can vary from person-to-person, and there is no absolute number that describes the correct location for each point. Overall, our results suggest that our system has the potential to guide users to locate their auscultation points using AR guidance, which can advance preventative care by enabling the general public to self-screen. In our evaluation of distance error, we considered subjects from five different races, including Asian, White, American Indian or Alaska Native, and Black or African American. However, our study did not find a significant relationship between race and distance error. This suggests that the performance of our system, in terms of distance measurement accuracy, is consistent across individuals from different racial backgrounds.

Figure 9a is the cumulative distribution function (CDF) on various age groups, it appears that the performance of ARSteth improves with increasing age group, with the 35-44 age group showing the best performance. There could be a variety of factors that lead to this outcome, including changes in chest wall structure as people get older [8]. We plan to explore this further in future work.Additionally, the improvement in performance with increasing age group could also be attributed to differences in the physical characteristics of the subjects, such as body mass index (BMI), which can affect the quality of sound transmission and detection. BMI was found to be a significant predictor of the quality of heart sound recordings [30]. As such, the older age group, which tends to have a higher BMI on average, may have more favorable physical characteristics for sound transmission and detection, leading to better performance.



(a) CDF of distance (error) against age groups.

(b) Time for identifying the auscultation points.

Figure 9: (Left)Distance error on different age groups. (Right)Time required for identifying the auscultation points.

5.1.3 *Timing.*

In this section, we evaluated the time required by users to position the stethoscope correctly on the auscultation points using ARSteth. This study measured the time taken from the start of movement until the stethoscope was placed in the correct position for each auscultation point. The results showed that the first auscultation point, i.e. Aortic, took the most time to locate, and this is due to the need to update and fine-tune the coarse estimation of the auscultation points along the trajectory of the movement at the beginning of the screening. However, in the next three target points, the calibration time was generally shorter because the coarse estimation had been improved while finding the Aortic point. On average, users were able to locate the other auscultation points in about 21.61 seconds when there was no projection enabled. However, the projection-based calibration of the auscultation point significantly reduced the time for users to find the auscultation points. The projection-based calibration ensured that the predicted points were closer to the actual target auscultation point, which resulted in less time spent on the initial auscultation point. The projection-based calibration reduced the time to locate each auscultation point to about 13.09 seconds. While it may take longer for users to locate the auscultation points using ARSteth compared to physicians, ARSteth requires less time to detect murmurs. During a standard physical examination, physicians normally listen to the heart for 30 to 60 seconds to determine the pace and rhythm of the heartbeats and to detect any abnormal sounds. On the other hand, ARSteth takes about one minute to identify all four auscultation points and "listen" at each point for 5 seconds.

5.1.4 *Usability.*

Table 1 compares the evaluation results of two different systems, ARSteth and a state-of-the-art intelligent stethoscope, the Eko 3M^{TM} Littmann® CORE Digital Stethoscope [7]. These evaluations are based on four factors: simplicity, guidance, confidence,

		Simplicity	**Guidance**	**Confidence**	**Aesthetics**
Group 1	ARSteth	4.1 (± 0.33)	**4.6 (± 0.49)**	4.6 (± 0.48)	4 (± 0.29)
	Eko	4.2 (± 0.41)	**3.7 (± 0.84)**	4.2 (± 0.41)	4.3 (± 0.47)
Group 2	ARSteth	4.3 (± 0.75)	4.5 (± 0.50)	**4.7 (± 0.44)**	4 (± 0.60)
	Eko	4.6 (± 0.48)	4.3 (± 0.75)	**3.6 (± 1.15)**	4.4 (± 0.48)

Table 1: Survey responses from two groups (averaged) and one standard deviation on four aspects of usability for ARSteth and Eko. Scores range from 1 (worst) to 5 (best).

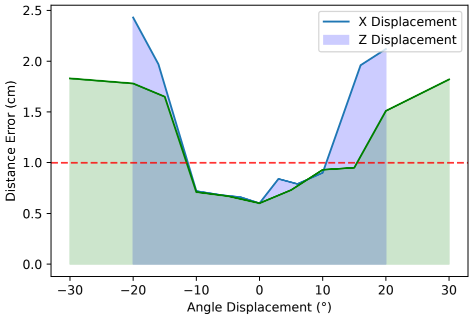


Figure 10: Error due to angle displacement. Blue (X) curve shows up and down displacement, green (Z) curve shows forward and backward displacement. The red dashed line represents the maximum error value ARSteth tolerates (1.0 cm).

and aesthetics. Simplicity refers to how easy the system is to use. Guidance measures the system's ability to direct users to the auscultation points, without necessarily assessing the correctness of the placement. Confidence refers to how users feel about the system's ability to correctly identify the auscultation points. A system that directs the stethoscope to an obviously incorrect location (e.g. outside of the chest area) will result in a lower confidence score. Aesthetics refers to the system's interface design. The test subjects are divided into two groups:

- *Group 1 (24 subjects)*, consisting of individuals with no prior experience with digital stethoscope technologies.
- *Group 2 (11 subjects)*, consisting of individuals with some familiarity with technology or prior experience with similar devices.

The results show that ARSteth provides more extensive guidance for users in Group 1, suggesting that ARSteth provides suitable guidance for individuals without any background. Furthermore, in Group 2, ARSteth scored significantly higher in confidence, which suggests that it may be more reliable in identifying auscultation points. Overall, Table 1 indicates that both ARSteth and Eko are relatively easy to use, but ARSteth provides additional guidance for users who may not be familiar with the technology. Additionally, ARSteth scored higher than Eko in terms of providing reliable results. Therefore, the results suggest that ARSteth may be a suitable option for the general public, particularly those who are not familiar with technology, as it can provide reliable self-screening at home.

5.2 Flexibility of Auscultation

In Section 5.1.2, ARSteth was found to provide reasonable estimates of auscultation points when users' sitting positions were restricted. However, maintaining an upright position for the duration of the examination can be difficult for users. Even though ARSteth accounts for rotations in the body through projection (Section 4.2), allowing users more flexibility while using the system is ideal. This evaluation aims to determine the maximum angle displacement in the x-axis and z-axis caused by sitting position variance while maintaining acceptable estimation of auscultation points.

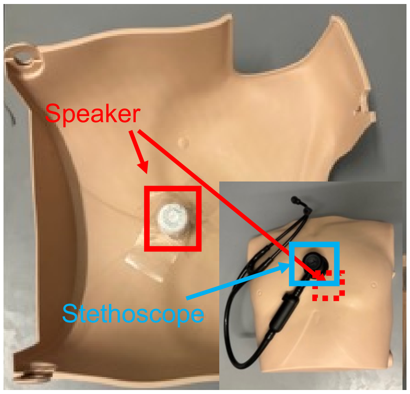


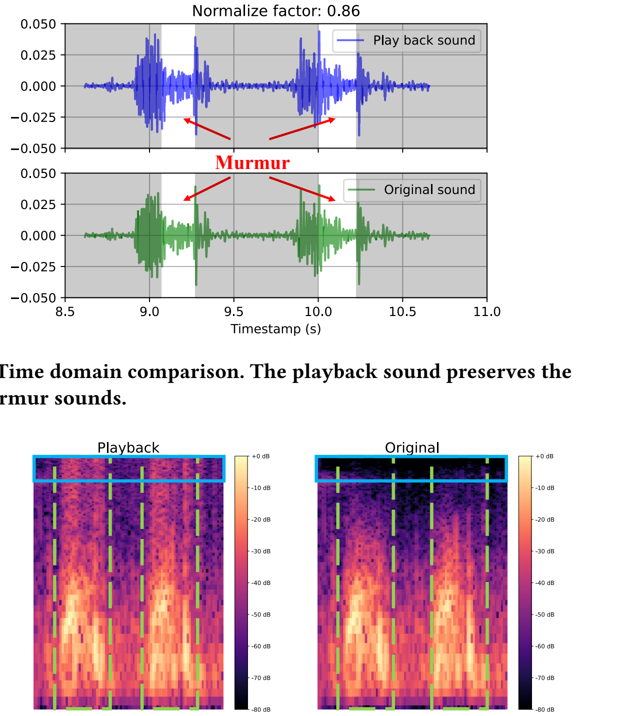
Figure 11: Speaker is placed inside of the manikin. Stethoscope is used to listen and record the sound on the other side.

The x-axis displacement is the angle between the line connecting the left and right shoulders and the horizontal line as shown in Figure 5a. The larger the difference in altitude between the shoulders, the greater the displacement in the x-axis. The z-axis displacement, illustrated in Figure 5b, increases if one shoulder moves towards the camera while the other moves farther away. In most cases, when users sit and use ARSteth for routine self-screening, there are typically three angles that describe their sitting position. Besides the two angles mentioned earlier, there is also a tilt angle of the upper body that should not affect the system since we model the auscultation point directly on the plane.

The objective of the following experiment is to test the maximum capability of the system in tolerating angle displacement in two directions, x-axis and z-axis. The subject is seated and the displacement angles are adjusted by moving the shoulders, with one angle being tested at a time. The maximum angle displacement is set at 20° for the x-axis and 30° for the z-axis. The experiment is performed on a male subject at the pulmonary auscultation point. The results show that the system can tolerate angle displacement in x-axis from about $-10°$ to $10°$ and in z-axis from about $-10°$ to $15°$. The maximum distance error is set to 1cm, which is an inflection point for both angle displacements. Beyond this threshold, the error increases rapidly. Therefore, the limit for both displacements is set to ±10° in the actual experiment. The experiment shows that angle displacement beyond these ranges leads to a much larger error in the distance.

5.3 Heart Murmur Detection

In Section 5.1, we evaluated ARSteth with subjects in normal health and no heart murmurs present. However, to evaluate ARSteth's ability to detect heart abnormalities, we need to simulate such abnormalities. Given that it is difficult to access patients with heart abnormalities, we used a silicone-made medical manikin to simulate a human body. We stripped the internal Nylon-made support, leaving only the 4.9 mm thick silicon skin, and attached a speaker (Momoho BTS0011) to the inner side of the skin to emulate the body



(a) Time domain comparison. The playback sound preserves the murmur sounds.

(b) Spectrogram of playback (left) and original (right) sounds. Two heart beats can be clearly identified in both figures.

Figure 12: Comparison between the playback sound and the original sound in both time domain and spectrogram.

surface, as shown in Figure 11. This is in line with existing techniques using heart sound recordings for telemedicine and physician training, which indicate that the errors introduced by the recordings can be tolerated. There are also existing works that use medical manikins to simulate heart sounds and perform auscultations in a lab environment [6, 24].

5.3.1 *Fidelity of Recorded Heart Sound.*

We first verified the fidelity of heart sounds recorded by the stethoscope by playing back sample heart sounds with heart murmurs collected with an Eko Stethoscope in the open dataset (Section 4.3.1). We continuously played back the sound with the speaker placed inside the manikin and then used the Eko stethoscope on the other side of the manikin at the same location. Figure 12 shows the original and playback sounds and their corresponding spectrogram. We used normalized cross-correlation to determine the time delay between the original and playback samples and scaled the recorded playback sound to eliminate the effect of intensity change due to the volume of the speaker. Although the playback sound introduced some high-frequency noise, the murmur was still present and identifiable in the playback sound. We confirmed this by passing the playback sound to the classifier we created in Section 12b, which successfully classified the sound. Despite the noise around 1 kHz, the model's ability to detect heart murmurs was unaffected because heart murmurs are typically found between 600 and 700 Hz.

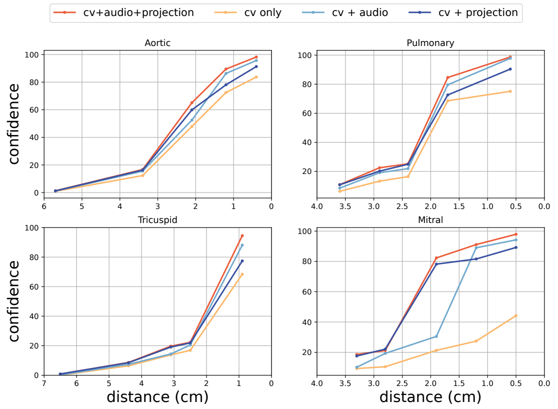


Figure 13: Comparison between four different methods of the detection confidence on an abnormal sound versus distance (error). The four methods are coarse estimation only (CV only), fine-tuned auscultation points (CV + audio), coarse estimation with projection calibration (CV + projection) and fine-tuning with projection calibration.

5.3.2 *Manikin-based Experiment.*

Our goal is to demonstrate how guidance can improve the quality of auscultation for better classification results. Specifically, we use the confidence of the model instead of the classification result to evaluate the effectiveness of our system. To simulate abnormal heart sounds as realistically as possible, we placed the speaker on the inner side of a manikin at the standard auscultation locations. We selected and looped samples with murmur sounds from the dataset that produced the highest confidence in the model output, and ran our system by following the steps provided by its guidance while using a custom-designed stethoscope on the manikin. To document the trajectory of the stethoscope, we marked the 4 auscultation points and sampled 5 locations along each trajectory.

Since multiple auscultation points surround some areas of a real body, sounds from different sources are combined into one, making it impossible to simulate accurately. To address this, we began moving the stethoscope from a location far away from the other three auscultation points, following standard procedures for how a physician would move the stethoscope, to the target auscultation, where we placed a speaker playing a recording at the end of the trajectory to simulate heart sounds. Along each trajectory to the target auscultation point, we sampled sounds from 5 locations.

The results of our experiment are presented in Table 2. The distance to the target auscultation point decreases as we follow the guidance of our system, and the confidence in classifying heart murmurs increases monotonically. These findings support the reliability of our system for self-screening. Note that the target referred to here is the point where the acoustic calibration algorithm terminates, rather than the actual location of the speaker.

	Point 1		**Point 2**		**Point 3**		**Point 4**		**Point 5**	
(cm / %)	Dist.	Conf.	Dist.	Conf.	Dist.	Conf.	Dist.	Conf.	Dist.	Conf.
Aortic	5.7	**1.3**	3.4	16.7	2.1	65.1	1.2	89.5	0.4	**98.2**
Pulmonary	3.6	**10.9**	2.9	22.5	2.4	25.2	1.7	84.6	0.6	**98.7**
Tricuspid	6.5	**0.8**	4.4	8.7	3.1	19.7	2.5	22.3	0.9	**94.5**
Mitral	3.3	**18.6**	2.8	21.2	1.9	82.3	1.2	91.1	0.5	**97.9**

Table 2: The distance (error) to the target auscultation point and its corresponding confidence on heart murmur.

To compare the performance of each component in ARSteth, we repeated the experiment with coarse estimation using computer vision only for coarse estimation, coarse estimation with projection, and fine-tuned position without projection. Figure 13 illustrates the results, which were obtained by collecting sounds at the same distances as listed in Table 2 and obtaining corresponding classification confidences. The blue line representing fine-tuned estimation by sound achieved the second-best result, indicating that the fine-tuned location obtained by analyzing the sound from the stethoscope can help increase the chance of placing the stethoscope closer to the target point. In particular, the projection was found to be useful when the stethoscope was far away from the source of sound.

6 DISCUSSION

6.1 Limitation on the Evaluation

The current evaluation methodology for ARSteth has several limitations that should be acknowledged. One major limitation is that the evaluation is conducted on manikins, which may not accurately reflect the actual conditions of human subjects. Manikins lack the variety of human chest characteristics, such as chest size, shape, and hair, which may impact the accuracy of murmur detection. In addition, manikins may not fully simulate the respiratory and circulatory systems, which may result in inaccurate representations of real-world auscultation conditions.

Moreover, the use of public datasets has its own set of limitations. Public datasets are usually recorded from a limited number of subjects with specific medical conditions, which may not represent the entire population. This can lead to biased evaluation results and limit the generalizability of the findings. Furthermore, public datasets often include inherent noise and artifacts.

Another limitation of the current evaluation methodology is the error introduced from the playback of the public dataset through the manikin. The use of a manikin as an intermediary introduces additional error due to the differences between the manikin and real human subjects. For example, the material of the manikin's chest may not be the same as that of a human chest, leading to variations in sound transmission and attenuation. Additionally, the use of a manikin may also introduce different frequency responses, which may affect the detection of certain types of murmurs.

Despite the limitations mentioned above, our evaluation has shown that combining public datasets and playback through a manikin can be an effective evaluation method for ARSteth. The normalization and cross-correlation of the recordings and playback demonstrate the fidelity of the playback process. However, it is important to note that the evaluation results obtained through this method may not fully represent the performance of the intelligent stethoscope during real subject tests.

In summary, the current evaluation methodology has limitations that should be acknowledged. The use of manikins and public datasets may not fully reflect real-world auscultation conditions, leading to biased and limited evaluation results. Additionally, the playback through a manikin introduces additional error due to the differences between the manikin and real human subjects. Further research is needed to develop more comprehensive evaluation methods that address these limitations and accurately reflect the performance of intelligent stethoscopes during real subject tests.

6.2 Acceptable accuracy and error tolerance

The probability density function in Figure 8 indicates that the Aortic, Pulmonary, Tricuspid, and Mitral regions have the highest probabilities at 12.6mm, 11.2mm, 13.1mm, and 16.7mm, respectively, across all genders. However, these numbers may not be easily understood by non-medical professionals, and there is no established range of acceptable auscultation points in the medical field. To address this, we consulted with two cardiologists who confirmed that physicians typically place the stethoscope at the four standard cardiac regions corresponding to the heart valves, and move it to identify the best position for hearing heart sounds. Based on this guidance, our collaborating physician validated the auscultation points found by our system as acceptable because they were in the range of the corresponding main heart valve areas and produced high-quality heart sounds. Note that the diameter of an auscultation point can vary from person to person. Furthermore, we plan to create a benchmark to assess the distance error between auscultation points identified by physicians in a diverse group of test subjects, which would further support our work.

7 CONCLUSION

In this paper, we introduce ARSteth, an AR-assisted home self-screening system for cardiopulmonary diseases with a custom-designed low-cost stethoscope. We implement ARSteth using augmented reality to provide the user with real-time feedback in locating auscultation points. To further enhance the precision of auscultation points without the presence of medical professionals, we present a method to fine-tune the locations by continuously analyzing the sound from the stethoscope. Our results show that ARSteth utilizing our custom-designed stethoscope can provide quick (about 13.09 seconds for each auscultation point) and accurate guidance (13.2mm distance error on average) with less restrictions on users' sitting positions compared with 24.7mm distance error in our baseline. We also demonstrate its ability to detect potential heart diseases by evaluating abnormal sounds with a medical manikin. While further testing is required to assess the system's accuracy and feasibility for clinical applications such as murmur detection, our preliminary results suggest that ARSteth has the potential to enhance early detection for the general public.

ACKNOWLEDGMENTS

This research was partially supported by the National Science Foundation under Grant Numbers CNS-1704899, CNS-1815274, CNS-1943396, CNS-1837022, and CMMI-2218809. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Columbia University, NSF, or the U.S. Government or any of its agencies.

REFERENCES

[1] Bekah Allen, Robert Molokie, and Thomas J. Royston. 2020. Early Detection of Acute Chest Syndrome Through Electronic Recording and Analysis of Auscultatory Percussion. *IEEE Journal of Translational Engineering in Health and Medicine* 8 (2020), 1–8. https://doi.org/10.1109/JTEHM.2020.3027802

[2] Paolo Bifulco, Fabio Narducci, Raffaele Vertucci, Pasquale Ambruosi, Mario Cesarelli, and Maria Romano. 2014. Telemedicine supported by Augmented Reality: an interactive guide for untrained people in performing an ECG test. *Biomedical engineering online* 13, 1 (2014), 1–16.

[3] Jack Carfagno. 2022. AI-powered Smart Stethoscope Revolutionizing Remote Medicine. https://www.docwirenews.com/docwire-pick/ai-powered-smart-stethoscope-revolutionizing-remote-medicine/. Accessed: 2022-010-25.

[4] Tao Chen, Xiaoran Fan, Yongjie Yang, and Longfei Shangguan. 2023. Towards Remote Auscultation with Commodity Earphones. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems* (Boston, Massachusetts) *(SenSys '22)*. Association for Computing Machinery, New York, NY, USA, 853–854. https://doi.org/10.1145/3560905.3568084

[5] Muhammad Enamul Hoque Chowdhury, Amith Khandakar, Khawla Alzoubi, Samar Mansoor, Anas Tahir, Mamun Bin Ibne Reaz, and Nasser Al-Emadi. 2019. Real-Time Smart-Digital Stethoscope System for Heart Diseases Monitoring. *Sensors* (2019). https://doi.org/10.3390/s19122781

[6] Sophia A. da Silva-Oolup, Dominic Giuliano, Brynne Stainsby, Joshua Thomas, and David Starmer. 2022. Evaluating the baseline auscultation abilities of second-year chiropractic students using simulated patients and high-fidelity manikin simulators: A pilot study. *The Journal of Chiropractic Education* (2022). https://doi.org/10.7899/jce-21-1

[7] Eko. 2022. Ekoscope. http://www.ekoscope.com/. Accessed: 2022-09-25.

[8] Michael Fiechter, Tobias A. Fuchs, Christian Hengstenberg, Catherine Gebhard, Julia Stehli, Bernd Klaeser, Barbara E. Stähli, Robert Manka, Robert Manka, Costantina Manes, Felix C. Tanner, Oliver Gaemperli, and Philipp A. Kaufmann. 2013. Age-related normal structural and functional ventricular values in cardiac function assessed by magnetic resonance. *BMC Medical Imaging* (2013). https://doi.org/10.1186/1471-2342-13-6

[9] Houman Ghaemmaghami, Nayyar Hussain, Khoa Tran, Aiden Carey, Shamile Hussain, Farhan Syed, Anthony J. Sinskey, Kevin O'Hashi, and John W. Sperling. 2017. Automatic segmentation and classification of cardiac cycles using deep learning and a wireless electronic stethoscope. *null* (2017). https://doi.org/10.1109/lsc.2017.8268180

[10] Tamer Ghanayim, Lior Lupu, Sivan Naveh, Noa Bachner-Hinenzon, Doron Adler, Salim Adawi, Shmuel Banai, and Avinoam Shiran. 2022. Artificial intelligence-based stethoscope for the diagnosis of aortic stenosis. *The American Journal of Medicine* (2022).

[11] Dezhi Hong, Ben Zhang, Qiang Li, Shahriar Nirjon, Robert Dickerson, Guobin Shen, Xiaofan Jiang, and John Stankovic. 2012. SEPTIMU: Continuous in-Situ Human Wellness Monitoring and Feedback Using Sensors Embedded in Earphones. In *Proceedings of the 11th International Conference on Information Processing in Sensor Networks* (Beijing, China) *(IPSN '12)*. Association for Computing Machinery, New York, NY, USA, 159–160. https://doi.org/10.1145/2185677.2185727

[12] Kaiyuan Hou, Yanchen Liu, Peter Wei, Chenye Yang, Hengjiu Kang, Stephen Xia, Teresa Spada, Andrew Rundle, and Xiaofan Jiang. 2022. A Low-Cost In-situ System for Continuous Multi-Person Fever Screening. In *2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. 15–27. https://doi.org/10.1109/IPSN54338.2022.00009

[13] Kaiyuan Hou, Stephen Xia, and Xiaofan Jiang. 2022. BuMA: Non-Intrusive Breathing Detection Using Microphone Array. In *Proceedings of the 1st ACM International Workshop on Intelligent Acoustic Systems and Applications* (Portland, OR, USA) *(IASA '22)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3539490.3539598

[14] Osama Ibrahim, Mathew Sibbald, Wojciech Szczeklik, and Wojciech Leśniak. [n.d.]. Heart Auscultation.

[15] Christine Laine. 2002. The annual physical examination: needless ritual or necessary routine? *Annals of Internal Medicine* 136, 9 (2002), 701–703.

[16] Jian Ping Li, Amin Ul Haq, Salah Ud Din, Jalaluddin Khan, Asif Khan, and Abdus Saboor. 2020. Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare. *IEEE Access* 8 (2020), 107562–107582. https://doi.org/10.1109/ACCESS.2020.3001149

[17] H Liang, Sakari Lukkarinen, and Iiro Hartimo. 1997. Heart sound segmentation algorithm based on heart sound envelogram. In *Computers in Cardiology 1997*. IEEE, 105–108.

[18] Chengyu Liu, David Springer, Qiao Li, Benjamin Moody, Ricardo Abad Juan, Francisco J Chorro, Francisco Castells, José Millet Roig, Ikaro Silva, Alistair E W Johnson, Zeeshan Syed, Samuel E Schmidt, Chrysa D Papadaniil, Leontios Hadjileontiadis, Hosein Naseri, Ali Moukadem, Alain Dieterlen, Christian Brandt, Hong Tang, Maryam Samieinasab, Mohammad Reza Samieinasab, Reza Sameni, Roger G Mark, and Gari D Clifford. 2016. An open access database for the evaluation of heart sound algorithms. *Physiological Measurement* 37, 12 (nov 2016), 2181. https://doi.org/10.1088/0967-3334/37/12/2181

[19] Yanchen Liu, Stephen Xia, Jingping Nie, Peter Wei, Zhan Shu, Jeffrey Andrew Chang, and Xiaofan Jiang. 2022. aiMSE: Toward an AI-Based Online Mental Status Examination. *IEEE Pervasive Computing* 21, 4 (2022), 46–54. https://doi.org/10.1109/MPRV.2022.3172419

[20] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, Matthias Grundmann, and Matthias Grundmann. 2019. MediaPipe: A Framework for Building Perception Pipelines. *arXiv: Distributed, Parallel, and Cluster Computing* (2019). https://doi.org/null

[21] Jingping Nie, Yanchen Liu, Yigong Hu, Yuanyuting Wang, Stephen Xia, Matthias Preindl, and Xiaofan Jiang. 2021. SPIDERS+: A light-weight, wireless, and low-cost glasses-based wearable platform for emotion sensing and bio-signal acquisition. *Pervasive and Mobile Computing* 75 (2021), 101424. https://doi.org/10.1016/j.pmcj.2021.101424

[22] Jingping Nie, Hanya Shao, Minghui Zhao, Stephen Xia, Matthias Preindl, and Xiaofan Jiang. 2022. Conversational AI Therapist for Daily Function Screening in Home Environments. In *Proceedings of the 1st ACM International Workshop on Intelligent Acoustic Systems and Applications* (Portland, OR, USA) *(IASA '22)*. Association for Computing Machinery, New York, NY, USA, 31–36. https://doi.org/10.1145/3539490.3539603

[23] Jingping Nie, Minghui Zhao, Stephen Xia, Xinghua Sun, Hanya Shao, Yuang Fan, Matthias Preindl, and Xiaofan Jiang. 2023. AI Therapist for Daily Functioning Assessment and Intervention Using Smart Home Devices. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems* (Boston, Massachusetts) *(SenSys '22)*. Association for Computing Machinery, New York, NY, USA, 764–765. https://doi.org/10.1145/3560905.3568074

[24] Antonia Quinn, Antonia Quinn, Jennifer Kaminsky, Jennifer Kaminsky, Jennifer Kaminsky, Andrew J. Adler, Shirley Eisner, Andrew Adler, Robin K. Ovitsh, Shirley Eisner, and Robin Ovitsh. 2019. Cardiac Auscultation Lab Using a Heart Sounds Auscultation Simulation Manikin. *MedEdPORTAL* (2019). https://doi.org/10.15766/mep_2374-8265.10839

[25] Luca Richeldi, Vincent Cottin, Gebhard Würtemberger, Michael Kreuter, Mariarosaria Calvello, and Giacomo Sgalla. 2019. Digital Lung auscultation: will early diagnosis of fibrotic interstitial lung disease become a reality? *American Journal of Respiratory and Critical Care Medicine* 200, 2 (2019), 261–263.

[26] Janice Hopkins Tanne. 2008. More than 26 000 Americans die each year because of lack of health insurance. *BMJ: British Medical Journal* 336, 7649 (2008), 855.

[27] Zeenat Tariq, Sayed Khushal Shah, and Yugyung Lee. 2019. Lung Disease Classification using Deep Convolutional Neural Network. In *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 732–735. https://doi.org/10.1109/BIBM47256.2019.8983071

[28] Ilias Tougui, Abdelilah Jilbab, and Jamal El Mhamdi. 2020. Heart disease classification using data mining tools and machine learning techniques. *Health technology* (2020). https://doi.org/10.1007/s12553-020-00438-1

[29] Sergio Varela-Santos and Patricia Melin. 2021. A new modular neural network approach with fuzzy response integration for lung disease classification based on multiple objective feature optimization in chest X-ray images. *Expert Systems with Applications* 168 (2021), 114361. https://doi.org/10.1016/j.eswa.2020.114361

[30] Nikolas Wanahita, Franz H. Messerli, Sripal Bangalore, Apoor S. Gami, Virend K. Somers, and Jonathan S. Steinberg. 2008. Atrial fibrillation and obesity—results of a meta-analysis. *American Heart Journal* (2008). https://doi.org/10.1016/j.ahj.2007.10.004

[31] Laura Whitney. 2022. Points of Auscultation | Anatomy Slices. https://3d4medical.com/blog/points-of-auscultation-anatomy. Accessed: 2022-09-11.

[32] Stephen Xia and Xiaofan Jiang. 2020. PAMS: Improving Privacy in Audio-Based Mobile Systems. In *Proceedings of the 2nd International Workshop on Challenges in Artificial Intelligence and Machine Learning for Internet of Things* (Virtual Event,

Japan) *(AIChallengeIoT '20)*. Association for Computing Machinery, New York, NY, USA, 41–47. https://doi.org/10.1145/3417313.3429383

[33] Stephen Xia and Xiaofan Jiang. 2022. AvA: An Adaptive Audio Filtering Architecture for Enhancing Mobile, Embedded, and Cyber-Physical Systems. In *2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. 118–131. https://doi.org/10.1109/IPSN54338.2022.00017

[34] Stephen Xia, Jingping Nie, and Xiaofan Jiang. 2021. CSafe: An Intelligent Audio Wearable Platform for Improving Construction Worker Safety in Urban Environments. In *Proceedings of the 20th International Conference on Information Processing in Sensor Networks (Co-Located with CPS-IoT Week 2021)* (Nashville, TN, USA) *(IPSN '21)*. Association for Computing Machinery, New York, NY, USA, 207–221. https://doi.org/10.1145/3412382.3458267

[35] Stuart J Youngner, Robert M Arnold, and Renie Schapiro. 2002. *The definition of death: contemporary controversies*. JHU Press.