# Informative Path Planning for Automatic Robotic Auscultation

**Poster** · May 2021



Yifan Zhu
University of Illinois, Urbana-Champaign

# Informative Path Planning for Automatic Robotic Auscultation

Yifan Zhu[1], Alexander Smith[2], and Kris Hauser[1]

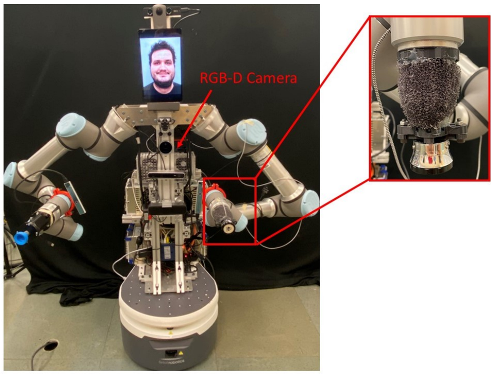


Fig. 1. The TRINA robot used in the experiments, with a An RGB-D camera (Intel Realsense L515) mounted on the torso to capture patient point clouds. A zoomed-in view of the digital stethoscope (ThinkLabs) is presented on the right.

## I. Introduction

Tele-medicine with robots has great potential to reduce risks to healthcare providers by minimizing person-to-person contact. This paper explores the task of auscultation, i.e., listening to heart and lung sounds via a stethoscope, which is an important routine physical screening procedure. In past work, tele-medicine systems have been built for performing auscultation and echocardiography [1, 2] through teleoperation, but automatic auscultation has yet not been explored. Automation can reduce operation time compared to tele-operation, enable physicians to perform procedures with low amounts of robot training, and reduces cognitive load and tedium. This project demonstrates the first automatic robotic system capable of performing heart and lung auscultation, and our experiments demonstrate that our system is capable of performing high-quality heart and lung auscultation in a small number of actions.

## II. Method

Suppose that sound quality estimators for heart and lung sounds are available, and denoted by $e_s(r)$, where $s$ is an anatomical structure and $r$ is a stethoscope recording, the goal is to generate high quality heart or lung sounds by placing a stethoscope at anatomically relevant portions of a patient's chest and back, while keeping the number of auscultation actions small. Letting $A$ be a region on the patient surface near structure $s$, we wish to solve the problem $\max_{x \in A} e_s(r(x))$ with a relatively small number of auscultation locations $x$. Our method consists of 2 phases. In the *offline phase*, we build a dataset of heart and lung stethoscope recordings of humans with labeled sound qualities and train heart and lung *sound quality estimators*. During the *online phase*, the robot first performs visual registration of the patient to a reference human model with labeled *clinical auscultation locations*. A prior of the *observation model* is constructed based on the visual registration, which encodes information on the locations of good sound qualities. Finally we use Bayesian Optimization (BO) to select locations to auscultate to find the best sounds utilizing both the learned sound quality estimators and the observation model.

### A. Bayesian Optimization Formulation

The BO approach estimates the quality of sound locations across the patient surface using a residual Gaussian Process (GP) model. The prior quality map is determined from visual registration (Section II-C), and then during auscultation the GP adapts the sound quality estimate to fit the observations.

For each anatomical location, the sound quality model is a sum of a prior mean function $\mu_{s,p}(x)$ and a GP residual function $f_s(x)$: $e_s(r(x)) \approx \mu_{s,p}(x) + f_s(x)$. Given observations $\bar{y} = (y_1, \ldots, y_k)$ at locations $\bar{x} = (x_1, \ldots, x_k)$ the standard GP equations predict a Gaussian distribution at a new evaluation location with mean $E[f_s(x)] = \mu_{\bar{y}}(x)$ and variance $Var[f_s(x)] = \sigma_{\bar{y}}(x)^2$. Since the GP models residuals with respect to the prior, we subtract the prior from the sound quality as the GP observations: $y_i = e_s(r(x_i)) - \mu_{s,p}(x_i)$. In each iteration of BO, an acquisition function is used to determine the next location to observe the data, and we use the popular upper confidence bound (UCB) function. UCB makes the choice

$$\arg\max_{x \in A} \mu_{s,p}(x) + \mu_{\bar{y}}(x) + \beta \sigma_{\bar{y}}(x), \tag{1}$$

where $\beta$ is an exploration weight that regulates how much bonus is given to uncertainty in the prediction. The maximization is performed in brute force fashion across locations sampled across $A$ (Algorithm 1).

### B. Sound Quality Estimator

We build a dataset of lung and heart stethoscope recordings to train our estimator. An expert grades each 10 s recording on a scale of 0 (worst) to 1 (best). We modify the method of Grooby et al. [3] for heart and lung quality classification and to handle our regression task. In particular, we apply the same features and adopt the TPOT automated machine learning algorithm for training. This achieves a 0.103 and 0.131 test MAE for heart and lung sounds, respectively.

### C. Visual Registration and Sound Quality Prior

We pre-define a set of locations on a patient body doctors typically auscultated as specific structures of the heart and

*This work was supported by NSF Grant #2025782

[1]: Y. Zhu and K. Hauser are with the Departments of Computer Science, University of Illinois at Urbana-Champaign, IL, USA. {yifan16, kkhauser}@illinois.edu

[2]: A. Smith is with the Carle Illinois College of Medicine, IL, USA. ads10@illnois.edu

**Algorithm 1:** Bayesian Optimization Auscultation

**Input:** Structure $s$ (heart/lung), prior $\mu_{s,p}$, exploration weight $\beta$, max iterations $N$;
Initialize $\bar{x} = \{\}$, $\bar{y} = \{\}$;
**for** *All regions $A$ near $s$* **do**
  **for** $k = 1, ..., N$ **do**
    $x_k \leftarrow \arg\max_{x \in A} \mu_{s,p}(x) + \mu_{\bar{y}}(x) + \beta\sigma_{\bar{y}}(x)$ ;
    **if** *Termination criteria met* **then**
      Go to next region;
    **else**
      Auscultate at $x_k$, obtain sound quality $e_k$;
      Set $\bar{x} \leftarrow \bar{x} \cup \{x_k\}$, $\bar{y} \leftarrow \bar{y} \cup \{e_k - \mu_{s,p}(x)\}$;
**return** Auscultation locations and recordings;

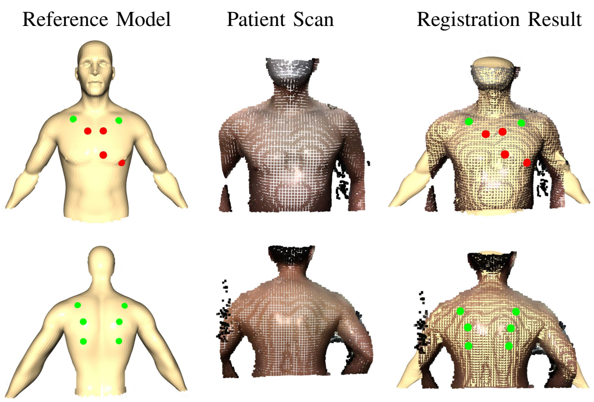


Fig. 2. From left to right are the labeled reference models, the patient scan point clouds, and the result of nonrigid registration. The auscultation locations for heart are labeled with red circles, and lung with green circles.

lung (Fig. 2). These locations are defined on a reference human mesh model. Visual registration then maps these points onto a point cloud of a patient captured with a RGBD camera. To do so, we first perform rigid registration with RANSAC followed by ICP to provide a good initial alignment, then nonrigid registration with nonrigid ICP proposed by Amberg et al. [4], where a local affine deformation model is used. The initial and final results of one case of visual registration is shown in Fig. 2. Currently, the accuracy of this procedure depends heavily on the patient's body shape.

The regions $A$ on the patient surface are limited to a certain distance from the estimated auscultation locations (Fig. 3, green circles). The prior model of sound qualities is defined by placing a negative exponential function at each of the registered auscultation locations in the projected coronal plane. We further apply a convolution with a 2D rectangular pulse function to obtain the resulting prior (Fig. 3).

## III. Results

The experiments use the TRINA robot, a mobile bimanual manipulator (Fig. 1). The Bayesian Optimization method

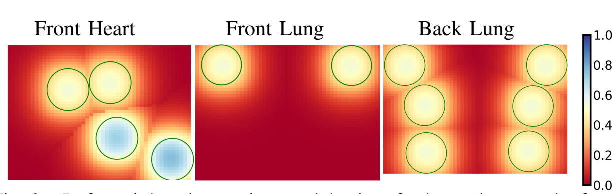


Fig. 3. Left to right: observation model priors for heart, lung on the front of the patient, and on the back. The regions $A$ are denoted with green circles.

TABLE I

Comparison between 3 auscultation methods

| Method | Reg. Time(s) | Total Time(s) | No. | Avg. Max |
|---|---|---|---|---|
| BO (heart) | 85 | 192 | 5 | 0.813 |
| RO (heart) | 85 | 168 | 4 | 0.750 |
| DT (heart) | 0 | 232 | 4 | 0.344 |
| BO (lung) | 171 | 785 | 34 | 0.516 |
| RO (lung) | 171 | 349 | 8 | 0.469 |
| DT (lung) | 0 | 417 | 8 | 0.391 |

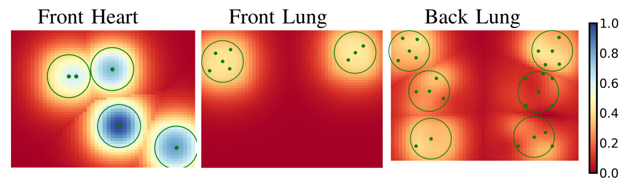


Fig. 4. Left to right: posterior sound quality predictions for heart, front lung, and back lung. Observed locations are marked as green dots.

(denoted as BO) is compared with 2 two baselines: registered locations only (RO) and teleoperation via a human operator (DT). In RO, the robot is controlled automatically to perform the registration step but only auscultate at registered locations. In DT, a trained human operator auscultates a patient using direct tele-operation controlled by a virtual reality headset and controllers (Oculus Quest 2) and is asked to auscultate at the locations specified in Fig. 2. The operator is a medical school student who has received auscultation and teleoperation training. The operator has feedback from the stethoscope during this procedure. In all of our experiments, we record a 10-second stethoscope audio clip at each auscultation location. The quality of each recording is graded by the operator in a follow-up phase, wherein recordings within each location were presented in randomized order.

The posterior observation model for BO is shown in Fig. 4. Table 1 compares time spent on visual registration (Reg. Time), total time spent auscultating (Total Time), number of auscultations (No.), and maximum quality found in each region, averaged over heart and lung regions (Avg. Max). BO is able to obtain higher quality sounds than RO, but spends more time auscultating. Both RO and BO obtain significantly higher quality sounds compared to DT. In addition, automatic auscultation saves time on execution compared to DT, except for BO for lung. However, the thoroughness of BO can be adjusted to reduce or limit the number of auscultations, at a possible sacrifice of quality.

## References

[1] M. Giuliani, D. Szczęśniak-Stańczyk, N. Mirnig, G. Stollnberger, M. Szyszko, B. Stańczyk, and M. Tscheligi, "User-centred design and evaluation of a tele-operated echocardiography robot," *Health and Technology*, vol. 10, no. 3, pp. 649–665, 2020.

[2] G. Yang, H. Lv, Z. Zhang, L. Yang, J. Deng, S. You, J. Du, and H. Yang, "Keep Healthcare Workers Safe: Application of Teleoperated Robot in Isolation Ward for COVID-19 Prevention and Control," *Chinese Journal of Mechanical Engineering*, vol. 33, no. 1, 2020.

[3] E. Grooby, J. He, J. Kiewsky, D. Fattahi, L. Zhou, A. King, A. Ramanathan, A. Malhotra, G. Dumont, and F. Marzbanrad, "Neonatal heart and lung sound quality assessment for robust heart and breathing rate estimation for telehealth applications," English, *IEEE Journal of Biomedical and Health Informatics*, 2020.

[4] B. Amberg, S. Romdhani, and T. Vetter, "Optimal Step Nonrigid ICP Algorithms for Surface Registration," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 8, IEEE, 2007, pp. 1–8.