2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Acoustic Collision Detection and Localization for Robot Manipulators

Xiaoran Fan, Daewon Lee, Yuan Chen, Colin Prepscius,
Volkan Isler, Larry Jackel, H. Sebastian Seung and Daniel Lee

***Abstract*— Collision detection is critical for safe robot operation in the presence of humans. Acoustic information originating from collisions between robots and objects provides opportunities for fast collision detection and localization; however, audio information from microphones on robot manipulators needs to be robustly differentiated from motors and external noise sources. In this paper, we present *Panotti*, the first system to efficiently detect and localize on-robot collisions using low-cost microphones. We present a novel algorithm that can localize the source of a collision with centimeter level accuracy and is also able to reject false detections using a robust spectral filtering scheme. Our method is scalable, easy to deploy, and enables safe and efficient control for robot manipulator applications. We implement and demonstrate a prototype that consists of 8 miniature microphones on a 7 degree of freedom (DOF) manipulator to validate our design. Extensive experiments show that *Panotti* realizes near perfect on-robot true positive collision detection rate with almost zero false detections even in high noise environments. In terms of accuracy, it achieves an average localization error of less than 3.8 cm under various experimental settings.**

I. Introduction

Collision detection is an important capability for manipulator operation in dynamic and uncertain environments. A robot colliding with the environment can damage itself or its surroundings. It can also cause harm to humans in the workspace. With the capability to detect collisions, the robot can stop. If the collisions can be localized, the robot can plan actions to avoid further impact. As a result, significant research has been dedicated to collision detection and localization for robot arms.

Existing approaches for collision detection and localization can be divided into two groups based on their sensing methods. The first group uses proprioception, *e.g.* motor torque, position, velocity and momentum readings coupled with inverse kinematics and dynamics [1]–[7]. As they are model-based techniques, the detection algorithms usually require accurate knowledge of the model parameters. In practice, however, it is challenging to obtain dynamic parameters due to noisy and time varying properties of motors. The second set of solutions rely on exteroceptive sensors such as cameras and tactile sensors [8]–[12]. The feedback from tactile sensors is typically limited to the area surrounding the sensor. Covering an entire arm with tactile sensors can be costly and difficult. Cameras, on the other hand, require line of sight to detect collisions. Obstacles in the workspace can cause occlusions which limit their effectiveness.

All authors are with the Samsung AI Center NY, 123 West 18th Street, New York, New York 10011

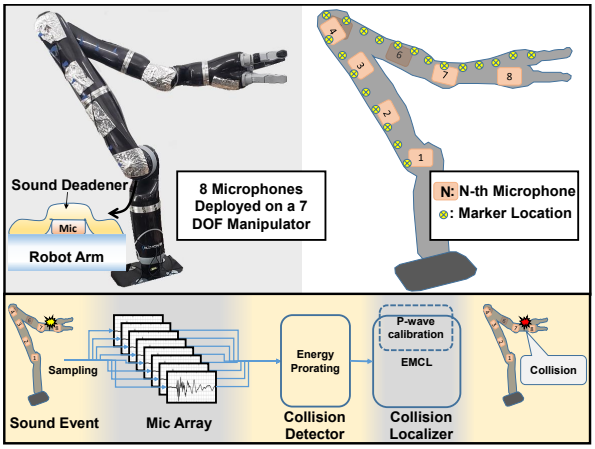

Fig. 1. *Panotti* system configuration and overview.

In this paper, we focus on a new modality, sound, and introduce a novel solution based on acoustic perception. Specifically, we present the *Panotti* collision detection and localization system for robot manipulators[1]. *Panotti* is the first system to use low-cost microphones deployed on the arm to detect and localize collisions.

In order to accurately detect collisions and localize them using microphones on a robot arm, two sets of challenges must be overcome. First, the system must distinguish between sounds caused by a collision and other sounds. This is difficult because there are many sound profiles which are similar to collisions. Such sounds can come from the outside and also from inside the robot (e.g. from the motors). In Section III-A, we present our approach for detecting collisions using a series of custom filters. The second challenge is regarding localization. Classical methods used in microphone and sonar arrays do not directly apply in our case. In particular, these methods usually rely on accurate detection of the onset time (the time for the sound to reach the microphone) from all microphones. In our case, we realized that while the onset time can be determined accurately for the microphone closest to the collision, it is hard to detect for the other microphones. However, those microphones still provide useful information which can be exploited for localization. In Section III-B, we present a novel method based on embedding the available information along with the knowledge about the calibrated microphone P-wave onset time onto an one-dimensional manifold in order

[1]The *Panotti* were a mythical group of people in Greek Mythology. They were described as possessing large ears that covered their entire bodies [13].

978-1-7281-6211-9/20/$31.00 ©2020 IEEE

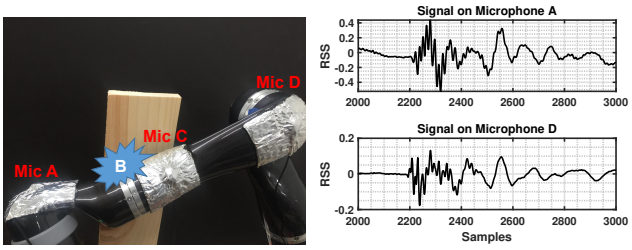

Fig. 2. **Left:** Three microphones deployed along a robot arm. A collision happens at location B. **Right:** The received signal strength (RSS) on microphones A and D under the this collision.

to localize the sound source.

To showcase these capabilities, we present a complete system where we deployed *Panotti* with eight microphones on a 7-DOF robot arm. In Section V, we present results from experiments where we show that *Panotti* can accurately detect collisions with an average localization error within 3.8 cm in challenging real-life settings. Our results establish *Panotti* as an accurate, low-cost collision detection and localization system for robot manipulators.

Our contributions can be summarized as follows:

- *Panotti* is the first system that detects and localizes robot collisions using low-cost microphones.
- We customize several signal processing algorithms to specifically address our unique design challenges. Moreover, we design a localization algorithm by exploiting the robot structure. The system requires very little calibration and runs in real-time.
- We implement a *Panotti* prototype on a 7 DOF manipulator with a multi-track microphone array, and conduct comprehensive evaluations. Our multi-scenario real world experiments take into consideration various over-the-air impulsive sound sources, on table and floor impulsive sound sources, multiple test objects colliding with the robot arm, and the moving robot arm colliding with multiple test objects. Our results show that *Panotti* realizes close to 100% on-robot collision true positive detection rate and close to 0% off-robot sound false positive detection rate, meanwhile achieving less than 3.8 cm average localization error.

In the next section (Section II), we introduce the physics behind collision on robots and present an overview of related work. The system design is detailed in Section III. An implementation (Section IV) and performance evaluation (Section V) then follows. Section VI concludes with a summary of our results and future research directions.

II. Background

In this section, we first provide a brief overview of the physics behind robot collisions in order to highlight the challenges in audio based approaches. Next, we provide a summary of related work in wireless localization.

A. Physics behind On-robot Collisions

Sound waves resulting from collisions that happen between robots and surrounding obstacles are of an impulsive nature. These impulse waves are lamb waves (plate waves) [14] because robots are generally constructed with solid plates. The ratio of wavelength to thickness of the robot plates is large. For example, if a 50 Hz wave propagates in the robot with a velocity of 1000 m/s, the wavelength is equal to 20 m, which is much larger than the thickness of robot plates (less than 1 cm).

Plate wave propagation in robots is of a dispersive nature, which causes different frequency components to have different velocities [15], [16]. A deployed example is shown in Figure 2. Three microphones are mounted along a robot arm at locations A, C and D. The collision happens at location B. Due to the dispersion, as illustrated in the right figure, each microphone will receive dissimilar signal with the same collision, which makes correlation based TDoA methods infeasible. Furthermore, even for the same frequency, plate wave propagation velocity depends on the robot structural and material characteristics, such as modulus of elasticity, density, Poisson's ratio, and slab thickness [17]. As illustrated in Figure 2, even with the on-robot distance $d_{AB} < d_{BD}$, due to the robot structural heterogeneity, the microphone at location D might receive the collision earlier than the microphone at location A. The heterogeneity consequently hinders the system determining the onset time or TDoA. Lastly, the relative location of each microphone changes as the robot arm moves around, making triangle-based localization algorithms not feasible.

B. Related Work in Wireless Localization

Wireless localization is of vital importance in several domains such as sonar [18], robotics [19], wireless sensor networks [20], and communication [21]. Existing methods can be categorized into three categories: i) signal strength fingerprinting, ii) beamforming based source localization, and iii) multilateration.

Localization through RSS signature fingerprinting is based on the wave (acoustic or radio frequency) propagation characteristics [22]. This approach requires a transmitter proactively sending known signals, and a RSS-location map is constructed by measuring the RSS at different distances from the signal source. Besides this approach being vulnerable to the ever changing wireless channel state information (multipath and fading), the signal source is randomly generated in each collision, which prevents us from using this fingerprinting approach. Beamforming approaches find the source location by shifting the phases of received signal copies until they find the maximum similarity in the spatial domain [23]. However, the beamforming approach assumes the wave propagates in a homogeneous medium. It also requires the receivers' geometric layout and wave velocity to be known, which is not feasible in the *Panotti* design.

Multilateration localizes the sound source based on the measurement of time of arrival (ToA) or time difference of arrival (TDoA) of waves across the receiver array [24]. Traditionally, this approach also requires the wave velocity and receivers' geometrical information. However, in the design of *Panotti*, we generalize the Multilateration method

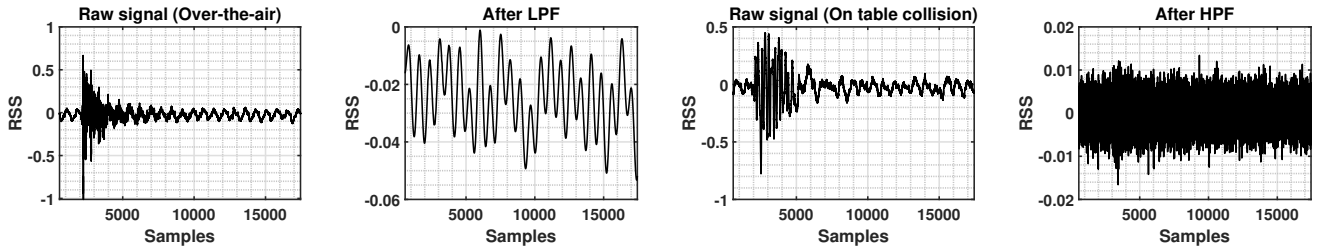

Fig. 3. From left to right: raw over the air signal, low-pass filtered over the air signal, raw on table signal, and high-pass filtered on table signal.

by calibrating the microphones' P-wave TDoA onto an one-dimensional manifold and then designing a scoring function to localize the sound source.

III. System Design

Panotti involves N=8 low-cost ($\sim$10$ each) miniature-microphones deployed along a 7 DOF robot arm. An important feature of *Panotti* is that the location information of microphones is not required.

In the rest of this section, we explain the details of each design component, including the on-robot collision detection method (§III-A), the collision localization algorithm, the epicenter multilateration for collision localization (*EMCL*) (§III-B), and the motor noise suppression mechanism in (§III-C).

A. On-robot Collision Detection

Similar to on-robot collisions, off-robot sounds can be impulsive by nature, and thus they can be captured by microphones and mistakenly considered as on robot collisions. Therefore, we need to carefully investigate the characteristics of off-robot sounds and detect on-robot impulses only.

In *Panotti*, the first thing we do is to physically isolate microphones from the robot's surroundings. As shown in Figure 1, in our implementation, we seal microphones with a sound deadening material [25]. The sound deadening material reduces the wide band over-the-air audio signal strength by approximately 20 dB. In this way, *Panotti* majorly receives signal generated on the robot and mediums coupled with the robot (*e.g.* the support table and floor).

However, as shown in Figure 3 left, the residual over-the-air signal still remains significant. Determining the on-robot collision by simply thresholding the raw signal is not robust and generalizable. In *Panotti* design, we apply a low pass filter (LPF) with very low cut of frequency ($<$ 200 Hz) to identify the over-the-air sound. The key intuition is on-robot collisions introduce micro vibrations along the robot body, which translates to significantly higher amount of infrasound energy. Consequently, true on-robot collisions can be easily detected after LPF. On the other hand, as shown in Figure 3 left, the filtered over-the-air sounds are usually under the noise floor or very weak.

Collisions on the table or floor that supports the robot may also vibrate the robot slightly, which consequently generate infrasound energy. However, in wave propagation, the high frequency components attenuates exponentially faster than the low frequency components [26]. Also, comparing with audio signals propagating in the air, solid materials such as wood and concrete floor absorb several magnitudes more energy in the whole spectrum [27]. As such, followed by the LPF, we apply another high pass filter (HPF) to identify the on table and on floor sound. An example result is shown in Figure 3 right. As can be seen, for the example on table collision, the high frequency components are well under the noise floor. Detailed cut off frequency investigations for both filters are presented in Section V-A.1

Algorithm 1 On-robot collision detection

```
function DETECTION(s_1, s_2...s_N)
    for i ← 1 to N do
        sL_i ← LPF (s_i)
        sH_i ← HPF (s_i)
        if (||sL_i||_inf) ≥ γ_1 & (||sH_i||_inf) ≥ γ_2 then
            return on-robot collision detected
        end if
    end for
end function
```

The implementation of on-robot collision detection is illustrated in Algorithm 1. N is the number of microphones. s_i is the received signal on i_{th} microphone. γ_1 and γ_2 are two preset thresholds. *Panotti* adapts this simple physical microphone tweak with a mixed filtering approach to identify on-robot collisions. The detection accuracy remains high even under extreme tests (4% false negative rate). Detailed evaluations of this mechanism are presented in Section V-B.1.

B. Epicenter Multilateration for Collision Localization (EMCL)

We detect the onset time t_i of i^{th} microphone by finding its first peak above an energy threshold $\gamma = 3\sigma$, where σ is the standard deviation of the received signal copy. $T = [t_1, t_2, ..., t_N]$ is the onset time set. The relative onset time (ROT) is defined as:

$$ROT = T - min(T). \tag{1}$$

As discussed in Section II, triangulation based localization methods and correlation based onset time detection methods are not suitable in our scenario. Let's rethink our application scenario, we are trying to localize sound sources generated on robot arms that usually around 1 meter long. Although the arm can move, the microphone locations on the robot are

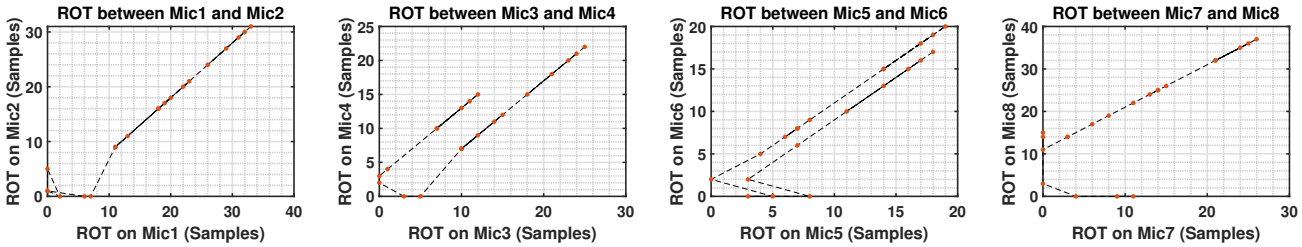

Fig. 4. From left to right: the 2 dimensional manifold projection in the microphone 1 and 2 space, the microphone 3 and 4 space, the microphone 5 and 6 space, and the microphone 7 and 8 space.

fixed, the arm's modulus of elasticity, density, Poisson's ratio, and slab thickness stay invariant during moving. For example in Figure 2, if two collisions happen at different locations between microphone A and B, $ROT(A) - ROT(B)$ will change but $ROT(D) - ROT(C)$ remains similar in two collisions. As such, when the collision location gradually moves between microphone i and microphone $i+1$, only the $ROT(i) - ROT(i+1)$ changes among microphones, *i.e. The ROT is an one dimensional manifold in the N dimension microphone space.* We can then calibrate and acquire the wave propagation properties beforehand, and apply them to localize the upcoming collision.

Overview of *EMCL*. In a manner analogous to locating the epicenter of earthquakes, we name our algorithm as Epicenter Multilateration for Collision Localization (*EMCL*). As shown in Figure 1, we first evenly mark along the robot arm at M locations, then we hit at each marked location clearly and derive the low dimensional manifold in the N dimension microphone space. When there is a new collision, we calculate the onset time for the strongest signal, and scoring each location along the calibrated manifold. We design the scoring function based on the manifold. Last, we conclude the collision location based on that score.

P-wave TDoA calibration. During the calibration, we have one ROT for each interested location along the robot arm, Hence the manifold $\mathcal{M}$ can be represented by a concatenation of M ROTs:

$$\mathcal{M} = [ROT_1; ROT_2; ...; ROT_M]. \tag{2}$$

The proposed localization method strongly depends on the estimation accuracy of manifold $\mathcal{M}$, which requires accurate onset times T. However, as illustrated in Figure 5, due to the dispersiveness and the attenuation in the signal, it is hard to determine accurate onset times for the microphones far from the signal source. Note the onset time difference between two adjacent microphones $\Delta t(i, i+1) = t_i - t_{i+1}$ is fixed if collisions didn't happen between microphone i and microphone $i+1$. According to this observation, we first obtain each $\Delta t(i, i+1)$ by proactively colliding near (but not in between) the microphone i or microphone $i+1$. Then we use this onset time difference $\Delta T = [\Delta t(1,2), \Delta t(2,3), ..., \Delta t(N-1, N)]$ to cleansing the manifold $\mathcal{M}$. The implementation is illustrated in Algorithm 2.

Algorithm 2 Cleansing the $M \times N$ dimension manifold

function CLEANSING($\mathcal{M}, \Delta T$)
  $V = NULL$
  for `i ← 1 to M` **do**
    $temp$ = ROT_i
    index = argmin(ROT_i)
    for `j ← index to 1` **do**
      if j≤2 **then**
        Break
      else
        $temp(j-2) = temp(j-1) + \Delta T(j-2)$
      end if
    end for
    for `j ← index to N` **do**
      if j≥N-1 **then**
        Break
      else
        $temp(j+2) = temp(j+1) + \Delta T(j+1)$
      end if
    end for
    V = Concatenate($V, temp$)
  end for
  return V
end function

Figure 4 describes the cleansed low dimensional (2D) projected manifold from the high dimensional ($N = 8$) microphone space. We evenly mark $M = 21$ locations (5cm spacing between two markers) along the robot arm in the calibration and colliding at each marker sequentially. As can be seen, there is a separable 1D manifold in each 2D projection. We call this calibration procedure as P-wave TDoA calibration due to the P-wave arrives fastest in earthquakes [28].

Localization. As an example shown in Figure 5, the first peaks of microphones near the collision are easy to find, however finding first peaks for the faraway microphones are difficult. In *Panotti* localization, we only find the onset time of first peak t_{ref} from the microphone that has the strongest signal s_i. Then we generate a set of *virtual onset time* $\mathcal{U}$ for the other microphones based on the manifold $\mathcal{M}$:

$$\mathcal{U} = \mathcal{M} - \mathbf{1}_N^T \otimes \mathcal{M}(i) + \mathbf{1}_M \cdot \mathbf{1}_N^T \cdot t_{ref}. \tag{3}$$

$\mathbf{1}_N$ is a column vector of 1 with N elements, $\mathcal{M}(i)$ is the i_{th} column in $\mathcal{M}$, and $\otimes$ is the Kronecker tensor

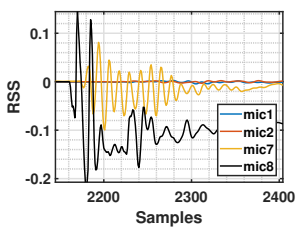

Fig. 5. Wave forms of 4 microphones. Collision happens close to mic 8. Mic 1 and 2 are far away.

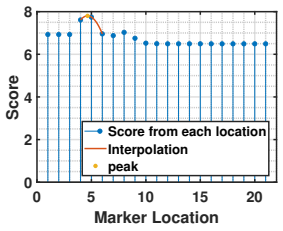

Fig. 6. Scoring and its interpolation. The collision location is determined as the peak of the fit quadratic curve.

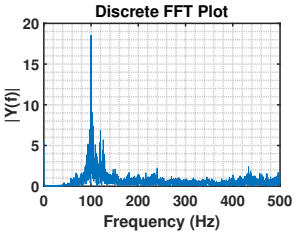

Fig. 7. Frequency domain of the recorded motor noise, there is a strong resonant frequency in 100Hz for the Kinova Jaco [30] manipulator.

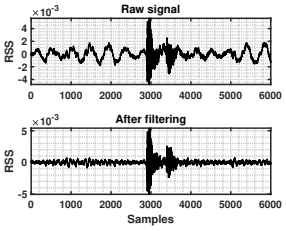

Fig. 8. Motor noise suppression. The upper plot is the raw signal with an arm moving collision, the lower plot shows the signal after filtering.

product [29]. The virtual onset time $\mathcal{U}$ is the shifted version of the original manifold $\mathcal{M}$ based on the incoming detected collision. The intuition here is we try to find the best matched marker location from this adjusted manifold $\mathcal{U}$. Next we adapt the virtual onset time $\mathcal{U}$ to derive the score for the i_{th} pre-determined marker location $\mathcal{S}(i)$:

$$\mathcal{S}(i) = \sum_{k=1}^{N} \mathcal{F}(\mathcal{U}_i(k))^2. \tag{4}$$

$\mathcal{F}$ is the scoring function and $\mathcal{U}_i$ is the i_{th} row in $\mathcal{U}$. In *Panotti* implementation, $\mathcal{F}(\mathcal{U}_i)$ is defined as the standard deviation within a small window w that expands each element in the virtual onset time $\mathcal{U}_i$. The physical intuition here is, when there is an upcoming collision, the signal starts to rise and oscillate in a faster rate if we found the *right* onset time. Larger disturbance translates to larger stand deviation value. As such, a higher score value represents to a higher chance to be the correct marker location. Moreover, we also investigate other scoring functions during *Panotti* design, detailed comparisons are discussed in section V-A.2.

Algorithm 3 Localization

```
function LOCALIZATION(S)
    k = argmax(S)
    if k == 1 then
        C = QuadraticFit([S(1), S(2), S(3)])
    else
        if k == N then
            C = QuadraticFit([S(k), S(k−1), S(k−2)])
        end if
        C = QuadraticFit([S(k−1), S(k), S(k+1)])
    end if
    return argmax(C)
end function
```

Finally, the localization method based on this score is illustrated in Algorithm 3. An example scoring and interpolation for a collision close to marker location 5 is illustrated in Figure 6. The final collision location is determined by looking for the peak index of the quadratic fit curve.

The elegance of *EMCL* is that it only requires the onset time for the microphone that has the strongest signal. Also, although different collision items generate different wave forms, the wave propagation speed should be similar across all frequency bins given fixed robot arm material and surface structure. It's the key for *Panotti* to robustly locate collisions with novel subjects. Detailed evaluations of the localization algorithm are presented in Section V-B.2.

C. Motor Noise Suppression

Motor noise is inevitable during robot operation. However, the majority of the noise comes from the nature frequencies of robot motors. As shown in Figure 7, the fundamental frequency of our Kinova Jaco is focused tightly around 100 Hz. We thus apply a band stop filter which has center frequency around 100 Hz onto the received signal to mitigate the motor noise. Figure 8 shows an illustration of the filtering result. As can be seen, the motor noise has been significantly mitigated. This process eventually brings *EMCL* with a better performance when the robot arm is moving. In practice, the arm's fundamental frequency can be acquired by calibration. For example, another robot arm Kinova Gen3 [31] we tested has the fundamental frequency around 80 Hz.

IV. Putting Together a *Panotti* Collision Detection System

We describe the system implementation in this section.

A. Testbed Setup

We deploy 8 Sujeetec miniature omnidirectional microphones [32] on a Kinova Jaco manipulator. The microphones are shielded with a Noico 80 mil automotive butyl and foil sound deadening material [25] and connected to a Zoom F8n MultiTrack Field Recorder [33]. The sampling rate Fs is set

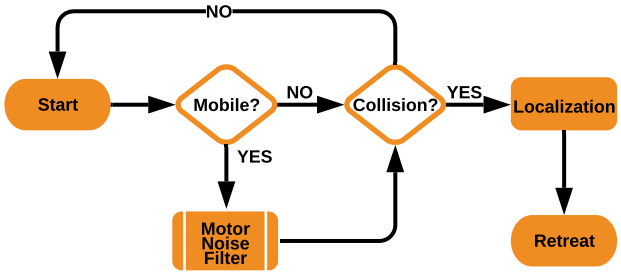

Fig. 9. System work flow of *Panotti* on the 7 DOF manipulator.

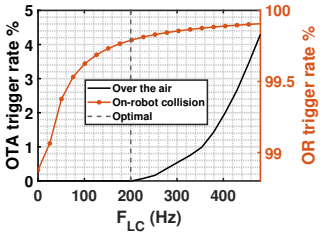

Fig. 10. OTA sound trigger rate and OR collision trigger rate vs. F_{LC}. Dashed line indicates the optimal F_{LC} selection.

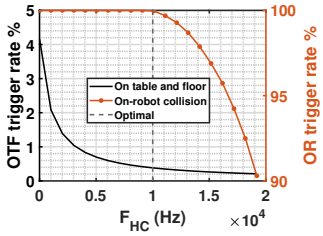

Fig. 11. OTF sound trigger rate and OR collision trigger rate vs. F_{HC}. Dashed line indicates the optimal F_{HC} selection.

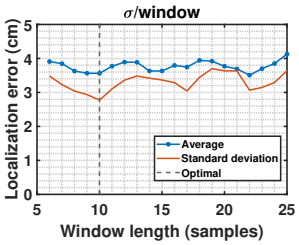

Fig. 12. Average localization error and σ vs. window length using standard deviation as the scoring function.

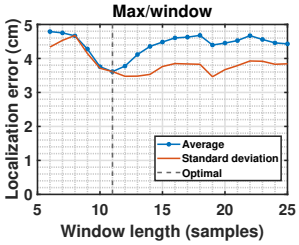

Fig. 13. Localization average error and σ vs. window length using max value as the scoring function.

as 48 KHz. Each microphone is driven by 24V phantom power provided by the Field Recorder. Signal processing is done by an Intel NUC7i7BNH computer [34].

The 8 microphones are evenly deployed along the 1 meter long Kinova Jaco arm. We evenly mark 21 calibration locations between microphone 1 and 8. The robot fundamental frequency is obtained by recording random movements of the robot arm for 240 seconds.

B. System Processing Flow

Figure 9 shows the processing flow of the system functioning modules. First, *Panotti* detects if the robot is in motion from the manipulation system call. Next, the motor noise filter is applied if the robot is in motion. Then the energy prorating mechanism is carried in the collision detection, and the *EMCL* localization algorithm follows if an on-robot collision is detected. Finally the robot starts to retreat according to the estimated collision location. We implement the robot retreat using Kinvoa SDK Ver. 1.5.1.

V. Experiments

We present the experiment and evaluation results in this section.

A. Micro-benchmark

We start with performing micro-benchmarks to evaluate system parameters in each function module. In this section, we aim to find the optimal system parameters by using a set of training data. We have collected 311 over-the-air (OTA) impulsive sound sources, such as human clasp, shout, and hitting use metal and wood etc. We have collected 276 on table and floor (OTF) sound sources, such as item drop on the table, human jumping, and hitting on the table/ground use a mallet. We also collected 286 on-robot (OR) collisions using a screw driver, which include 105 collisions (5 collisions at each marker location) for the P-wave TDoA calibration.

1) On-robot Collision Detection: Experiments in this section aim to i) understand the relationship between the performance of proposed on-robot collision detection rate and ii) finding the optimal filter settings of the energy prorating method.

Cut off frequency F_{LC}. We evaluate the trigger (detected as an OR collision) rate for OTA sound sources and OR collisions in different LPF cut off frequencies F_{LC}. As shown in Figure 10, the OTA trigger rate stays close to 0 percentage before 200 Hz while the OR trigger rate increases to around 99.7% at 200 Hz. We thus choose 200 Hz as F_{LC} in the *Panotti* implementation.

Scoring method	Absolute value	Absolute value^2	Max/window	Average/window	σ/window
Average/σ (cm)	4.38/3.87	4.42/3.76	3.87/3.12	4.28/3.57	3.47/2.79

TABLE I. Localization error and standard deviation (σ) for different scoring function..

Cut off frequency F_{HC}. Figure 11 shows the results when we evaluate the trigger rate for OTF sound sources and OR collisions in different HPF cut off frequencies F_{HC}. As can be seen, the OR trigger rate stays close to 100% before 10 KHz and starting to decreases when after 10 KHz. The OTF trigger rate start dropping sharply before 10 KHz and staying lower than 0.4% after 10 KHz. We thus choose 10 KHz as F_{HC} in the *Panotti* implementation.

2) Scoring function $\mathcal{F}$: Experiments in this section aim to i) validate our scoring method and ii) identify an optimal parameter for the scoring method's window.

Methods for the scoring function $\mathcal{F}$. As described in Section III-B, we use the standard deviation (σ) in a window as the scoring function and describe the detailed physical intuitions of this method. However, we also conducted experiments on other methods. We evaluate the localization performance for the following scoring methods: (1) directly using the signal's absolute RSS value exactly at the virtual onset time, (2) the square of the absolute RSS value, (3) the maximum RSS value within a small window (10 samples) around the virtual onset time, (4) the average RSS value within a small window around the virtual onset time, and (5) the standard deviation σ within a small window around

Fig. 14. Collision objects in our field study.

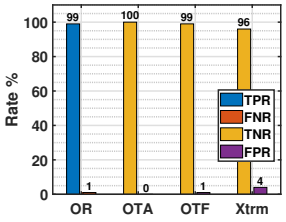

Fig. 15. Detection rate in various sound sources.

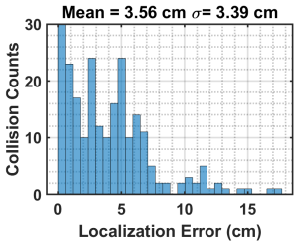

Fig. 16. Histogram of localization error when a screw driver collides with the robot.

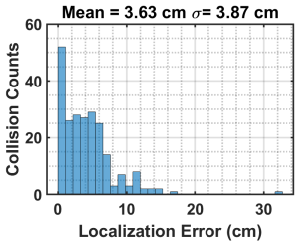

Fig. 17. Histogram of localization error with augmented noise when a screw driver collides with the robot.

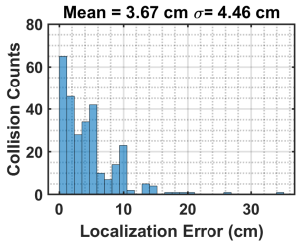

Fig. 18. Histogram of localization error when test objects multiple collides with the robot.

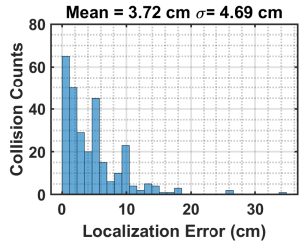

Fig. 19. Histogram of localization error with augmented noise when multiple test objects collides with the robot.

the virtual onset time. As shown in Table I, max value method (3) and standard deviation method (5) have the best performances. We evaluate the window length of these two methods next.

Window length for the scoring function $\mathcal{F}$. We vary the window length from 5 to 25 samples and try to find the optimal window length. As shown in Figures 12 and 13, these two methods are not very sensitive to the window length. We choose 10 samples which gives the best performance for standard deviation method as the window length, and proceed using this method in our field study next.

B. Field Study

After selecting optimal system parameters and obtaining the manifold from the training data in the micro-benchmark, we next conduct field study to evaluate *Panotti* in various scenarios. We collected more than 1000 real world on-robot collisions and off-robot sound sources. In particular, we collected 359 OTA impulsive sound sources, 224 OTF sound sources and 527 OR collisions for collision detection evaluation. In order to evaluate the localization accuracy, we collected another 241 OR collisions using a screw driver. Further, as shown in Figure 14, we randomly choose 16 items ranging from light weight plastic chip to wooden plank, human, and metal pole as the collision test objects. In total we have another 286 collisions using test objects. Lastly, we collected 137 collisions with test objects happened while the robot arm was actually moving.

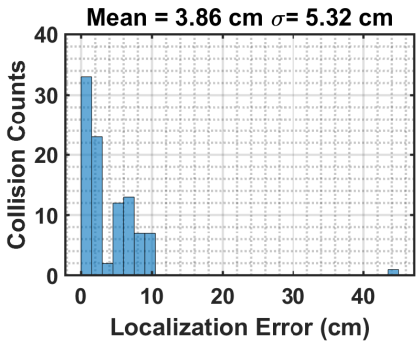

Fig. 20. Histogram of localization error when a moving robot collides with multiple test objects.

1) Collision Detection: In order to evaluate the performance of collision detection, we evaluate the true positive rate (TPR, succeed in detecting on-robot collisions), false positive rate (FNR, failure to detect on-robot collisions), true negative rate (TNR, succeed in detecting off-robot collisions), and false positive rate (FPR, failure to detect off-robot collisions) based on our field experiments. As shown in Figure 15, the TPR and TNR are close to 100% while the FPR and FNR are close to 0 percentage. Further, we collected 176 more OTA sound sources in extreme experiment settings, where the sound happens very close (< 0.3 m) to the robot arm. The result is still reasonably good with 96% TNR.

2) Localization Accuracy: We next evaluate the localization accuracy of *Panotti* in various experiment settings. We also study the localization accuracy by adding the robot arm motor noise onto the original recordings (augmented noise). We program the robot arm to move to multiple random locations and collect the motor noise. Lastly, we present the collision localization accuracy with a moving robot arm.

Collisions with a screw driver. The robot arm collides with a screw driver in this experiment. Figure 16 shows the histogram of the localization error. The error average is 3.56 cm and standard deviation is 3.39 cm. Most of the localization errors are small, and 95% of the localization errors are smaller than 10.71 cm. Figure 17 shows the histogram of the localization error with augmented motor noise. As can be seen, the average localization error stays similar but the standard deviation increases by 0.48 cm.

Collisions with test objects. The robot arm collides with test objects and human listed in Figure 14. Figure 18 and 19 show histograms of localization error with and without the augmented motor noise. Average localization errors are similar as colliding with the screw driver, and the standard deviations are increased around 1 cm. 95% of the localization errors are smaller than 11.24 cm and 11.57 cm with and without the augmented noise.

Collisions with test objects and robot moving. In this experiment, we program the robot arm, let it move randomly and collide with test objects. As shown in Figure 20, the average and standard deviation of localization error are 3.82 cm and 5.38cm. Except very small amount of outliers, the localization still remains accurate.

VI. Conclusion

In this paper, we introduced a robot collision detection and localization system using low cost microphones. We presented the design, implementation, and evaluation of *Panotti* in detail. We implemented our proposed system by deploying 8 low-cost microphones on a 7 DOF manipulator. Extended field experiments show *Panotti* realizes close to 100% on-robot collision true positive detection rate and close to 0% off-robot sound false positive detection rate, meanwhile achieving less than 3.8 cm average localization error. Moving forward, we plan to pursue subsequent real world experiments by deploying *Panotti* at actual warehouses for further validations.

There are multiple avenues for future research:

Colliding on fingers while grasping. In our current implementation, we pause the collision detection system when the robot is grasping items, because all our grasping tasks require the robot move to the target position then start grasping. However, in more general cases, robot might grasp and move the arm at the same time. Note that *Panotti* potentially enables fine-grained localization on the robot finger, which could be adapted for differentiating intentional grasping and collisions on the robot finger.

Alternative microphone deployments. These include optimizing the type, number and placement of microphones to further improve *Panotti*.

Localize multiple collisions. Multiple collisions can happen within the same time window in rare cases. A potential solution is to group microphones into multiple subsets and run our method alternatively.

Light collisions. Although most collisions are fatal and very strong, there are cases very light collision might happen when the robot collides with soft items. Light collision detection is very challenging due to lack of information. Also the definition of light collision is obscure. Nevertheless, it is desirable to investigate the limitations of *Panotti*.

Beyond robot arms. *Panotti* is implemented on a 7 DOF robot manipulators in this paper. It is however easy to adapt this system to other moderate size robots.

References

[1] E. Magrini, F. Flacco, and A. De Luca, "Estimation of contact forces using a virtual force sensor," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 2126–2133.

[2] S. Haddadin, A. De Luca, and A. Albu-Schäffer, "Robot collisions: A survey on detection, isolation, and identification," *IEEE Transactions on Robotics*, vol. 33, no. 6, pp. 1292–1312, 2017.

[3] G. Buondonno and A. De Luca, "Combining real and virtual sensors for measuring interaction forces and moments acting on a robot," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 794–800.

[4] J. Bimbo, C. Pacchierotti, N. Tsagarakis, and D. Prattichizzo, "Collision detection and isolation on a robot using joint torque sensing," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 7604–7609.

[5] L. Han, W. Xu, B. Li, and P. Kang, "Collision detection and coordinated compliance control for a dual-arm robot without force/torque sensing based on momentum observer," *IEEE/ASME Transactions on Mechatronics*, vol. 24, no. 5, pp. 2261–2272, 2019.

[6] Y. Lou, J. Wei, and S. Song, "Design and optimization of a joint torque sensor for robot collision detection," *IEEE Sensors Journal*, vol. 19, no. 16, pp. 6618–6627, 2019.

[7] A. De Luca, A. Albu-Schaffer, S. Haddadin, and G. Hirzinger, "Collision detection and safe reaction with the dlr-iii lightweight manipulator arm," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2006, pp. 1623–1630.

[8] P. M. Grice, M. D. Killpack, A. Jain, S. Vaish, J. Hawke, and C. C. Kemp, "Whole-arm tactile sensing for beneficial and acceptable contact during robotic assistance," in *2013 IEEE 13th International Conference on Rehabilitation Robotics (ICORR)*. IEEE, 2013, pp. 1–8.

[9] A. Jain, M. D. Killpack, A. Edsinger, and C. C. Kemp, "Reaching in clutter with whole-arm tactile sensing," *The International Journal of Robotics Research*, vol. 32, no. 4, pp. 458–482, 2013.

[10] A. Cirillo, F. Ficuciello, C. Natale, S. Pirozzi, and L. Villani, "A conformable force/tactile skin for physical human–robot interaction," *IEEE Robotics and Automation Letters*, vol. 1, no. 1, pp. 41–48, 2015.

[11] D. M. Ebert and D. D. Henrich, "Safe human-robot-cooperation: Image-based collision detection for industrial robots," in *IEEE/RSJ international conference on intelligent robots and systems*, vol. 2. IEEE, 2002, pp. 1826–1831.

[12] F. Flacco, T. Kröger, A. De Luca, and O. Khatib, "A depth space approach to human-robot collision avoidance," in *2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012, pp. 338–345.

[13] "Panotti," Webpage, 2017.

[14] I. Viktrov, "Rayleigh and lamb waves: physical theory and applications," *Chapter II*, 1967.

[15] Y. H. Lee and T. Oh, "The simple lamb wave analysis to characterize concrete wide beams by the practical masw test," *Materials*, vol. 9, no. 6, p. 437, 2016.

[16] K. Worden, "Rayleigh and lamb waves-basic principles," *Strain*, vol. 37, no. 4, pp. 167–172, 2001.

[17] Y. H. Lee and T. Oh, "The measurement of p-, s-, and r-wave velocities to evaluate the condition of reinforced and prestressed concrete slabs," *Advances in Materials Science and Engineering*, vol. 2016, 2016.

[18] S. Coraluppi, "Multistatic sonar localization," *IEEE Journal of Oceanic Engineering*, vol. 31, no. 4, pp. 964–974, 2006.

[19] J.-M. Valin, F. Michaud, J. Rouat, and D. Létourneau, "Robust sound source localization using a microphone array on a mobile robot," in *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453)*, vol. 2. IEEE, 2003, pp. 1228–1233.

[20] S. Li, X. Fan, Y. Zhang, W. Trappe, J. Lindqvist, and R. E. Howard, "Auto++ detecting cars using embedded microphones in real-time," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–20, 2017.

[21] J. J. Caffery Jr, "Wireless location in cdma cellular radio systems," Ph.D. dissertation, Georgia Institute of Technology, 1998.

[22] S. P. Tarzia, P. A. Dinda, R. P. Dick, and G. Memik, "Indoor localization without infrastructure using the acoustic background spectrum," in *Proceedings of the 9th international conference on Mobile systems, applications, and services*, 2011, pp. 155–168.

[23] X. Fan, H. Ding, Y. Zhang, W. Trappe, Z. Han, and R. Howard, "Distributed beamforming based wireless power transfer: Analysis and realization," *Tsinghua Science and Technology*, vol. 25, no. 6, pp. 758–775, 2020.

[24] Y. Zhou, J. Li, and L. Lamont, "Multilateration localization in the presence of anchor location uncertainties," in *2012 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2012, pp. 309–314.

[25] "Noico 80 mil sound deadner," Webpage, 2020.

[26] "Free-space path loss," Webpage, 2020.

[27] J. Achenbach, *Wave propagation in elastic solids*. Elsevier, 2012.

[28] "Seismic wave," Webpage, 2020.

[29] "Kronecker tensor product," Webpage, 2020.

[30] "KINOVA JACO Prosthetic robotic arm," Webpage, 2018.

[31] "KINOVA Gen3 Ultra lightweight robot," Webpage, 2018.

[32] "POM-3535L-2-R microphone driver," Webpage, 2018.

[33] "Zoom F8n MultiTrack Field Recorder," Webpage, 2015.

[34] "Intel NUC7i7BNH," Webpage, 2019.