Sound Source Tracking as a Frontier Exploration Heuristic for Search and Rescue using a Quadrupedal Robot

Francesco Marrato[1], M. Riley Cooper[2], and Joshua A. Marshall[3]

***Abstract*— This paper describes a system that integrates autonomous exploration agents for disasters scenarios with robots that have the ability to interpret their surroundings through sound. We present the development of an autonomous frontier exploration system that uses sound source tracking of human voices as a heuristic for search. Our system is capable of exploring previously unknown environments and locating a target sound source which emits human speech. We propose continuous and discrete approaches to sound source tracking and establish that measurements taken while the robot is stationary do not translate to measurable performance improvements. Our system was implemented on a Boston Dynamics Spot quadrupedal robot equipped with a four-microphone array. Field trials were conducted in a laboratory environment and in simulated disaster scenarios in both natural forest and urban environments. Across 64 trials, a success rate of 69 % and a success weighted by path length score of 0.42 was achieved. This research stands to demonstrate that sound source tracking tuned for the human voice is a suited heuristic to bring autonomous search agents used in disaster scenarios closer in capability to the humans they assist.**

I. INTRODUCTION

Since 2001, disaster robots have been successfully deployed by the Center for Robot-Assisted Search and Rescue (CRASAR) [10]. Off the shelf robots, such as the Inuktun MicroVGTV and VGTV Xtreme have been the go to, assisting in the search for disaster victims at the World Trade Center (2001) and the La Conchita Mudslide (2005) to name a few [10]. By introducing robots into their toolkit, rescue teams have increased worker safety, heightened situational awareness, and extended their reach to areas that were once inaccessible. Current disaster robots are teleoperated, relying on a team of workers to act as both the robots decision making systems and perform the stressful work of searching for disaster victims. Unfortunately, the split focus of operators and their limited ability to perceive the robot's surroundings through a computer screen have caused high failure rates in the field [10, Chapter 2].

[1]Francesco Marrato is with the Department of Electrical & Computer Engineering and the Ingenuity Labs Research Institute, Queen's University, Kingston Ontario, Canada 15fram@queensu.ca

[2]Riley Cooper is with the Department of Electrical & Computer Engineering and the Ingenuity Labs Research Institute, Queen's University, Kingston Ontario, Canada

[3]Joshua A. Marshall is with the Department of Electrical & Computer Engineering and the Ingenuity Labs Research Institute, Queen's University, Kingston Ontario, Canada joshua.marshall@queensu.ca

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada and the Ingenuity Labs Research Institute.

The codebase for this research is available at https://github.com/offroad-robotics/sst_as_a_heuristic_for_frontier_exploration

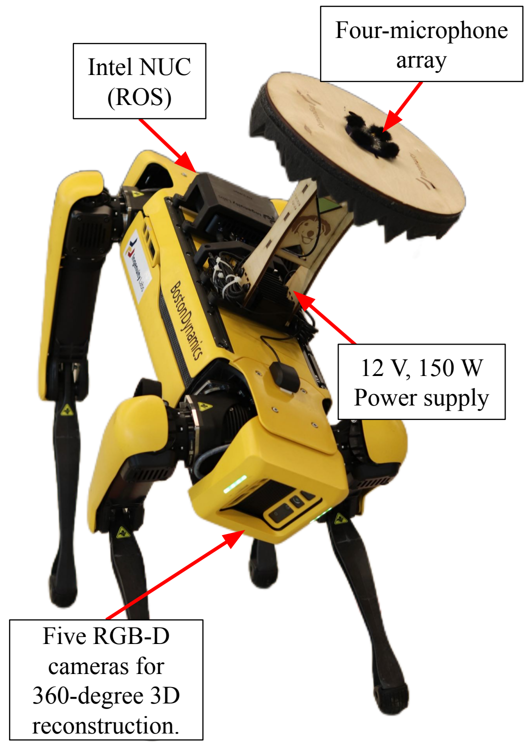

Fig. 1. Boston Dynamics Spot quadruped mobile robot outfitted with a four-microphone array (i.e., the "mushroom") as well as supporting visual sensing, power, and computing hardware.

Experimental disaster robots have been working to solve these issues. For example, NASA/JPL's Au-Spot project can autonomously map multi-kilometer underground environments, deciding where to search next by balancing coverage and travel costs [3]. The Honda Research Institute introduced robotic audition (robotic hearing) to a search and rescue Unmanned Aerial Vehicles (UAV), heightening the perception of its operators by visualizing the direction in which the UAV "hears" safety whistles [11].

However, the instinctual ability of rescue workers to locate disaster victims through calls for help is so far unmatched by today's search and rescue robots. In this paper, we study the use of robotic audition tuned specifically for human voices as a heuristic for autonomous exploration in search and rescue scenarios. We develop a system for the Spot quadrupedal robot that autonomously explores previously unseen environments and prioritizes search in the direction of human voices. We aim to demonstrate a reliable autonomous exploration system that uses human-like hearing to prioritize the search for disaster victims. Finally, we evaluate the system's performance in simulated disaster scenarios.

Briefly put, this paper makes the following contributions:

1) We explore the use of sound source tracking (SST) as a heuristic for frontier exploration in disaster robotics for human voice prioritized search.
2) We present an autonomous exploration system capable of finding a sound source target in simulated natural and urban disaster scenarios.
3) We investigate continuous and discrete approaches to SST to study the effects of local noise on robotic audition systems.

II. RELATED WORK

The development of a prototype search and rescue system that uses human voices to guide its exploration lies at the intersection of three areas of research. The documentation of current search and rescue robotics, the development of autonomous search strategies, and the use of audition in robotics all combine in development of this system.

The deployment of robots for search and rescue has been prevalent since the attacks on the World Trade Center in September of 2001. Accounts of robotic deployments carried out by R. Murphy's research team at Ground Zero are summarized by Davids in [4], which recognized the beginning of an emerging field. Murphy has documented robotics in search and rescue, summarizing all documented deployments of unmanned air, land, and sea vehicles between 2001 and 2014 in works [9] [10]. The term "disaster robot" was established in [10], recognizing that robots used in disaster events have a span of purposes and are not always optimized for the search and rescue of victims. The work in Murphy's [10, Chapter 6] establishes four types of field work for proper evaluation of emerging technologies in disaster robotics. Murphy also highlighted computer-mediated control as a limiting factor for successful deployments, establishing reliable autonomous systems as an open research question [10, Chapter 2].

The frontier-based approach to autonomous exploration has undergone evolution since its proposal by Yamauchi in [16]. Batinovic et al. [2] proposed the reduction of frontier candidates through clustering, proving an efficiency increase with no loss of environment coverage. The flexibility of the frontier selection heuristic has allowed search behaviors to be tailored to different applications. Originally, Yamauchi proposed that navigation should be conducted to the nearest accessible unvisited frontier [16], which has since been expanded. Bouman et al. [3] applied frontier exploration to a heavily equipped Spot robot, where the exploration policy is to maximize the covered area by navigating to frontiers with the highest potential to expand the robot's known world.

Proper evaluation of autonomous exploration is just as important as its development. Anderson et al. [1] proposed the Success weighted by Path Length (SPL) evaluation measure. SPL has allowed many researchers, such as Park et al. [15], to evaluate robotic exploration based on the efficiency of the path chosen. SPL helps to move field robotics past the holistic approach of "did it succeed or fail."

Robotic audition has been separately used for both victim localization in disaster robotics and for autonomous search. The inclusion of audition in robotics has come from the major contributions of Nakadai et al. in [14] and Grondin et al. in [7]. Both Nakadai's HARK [13] and Grondin's ODAS [6] have accelerated the adoption of audition in robotics by providing tools for sound source localization, tracking, and separation. Search and rescue UAV's have successfully implemented HARK to provide rescue workers with the direction of emergency whistles in simulated disasters [11]. Table-top exercises have also shown success for audition based autonomous search. Luo et al. used time difference of arrival to estimate the direction of a buzzer as the guiding direction in the TangentBug navigation algorithm [8].

In this paper, we approach the shortcomings of current disaster robots by developing a system that combines established research in autonomous exploration with robotic audition. We investigate SST as a capable heuristic for autonomous exploration in search and rescue scenarios.

III. FRONTIER EXPLORATION USING SST

This research utilized a Boston Dynamics Spot quadrupedal mobile robot as the base platform, selected for its capabilities in difficult terrain. Spot was outfitted with an Intel NUC11TNHi7 for compatibility with the Robot Operating System (ROS Noetic) and a four-microphone planar array for SST (Fig. 1).

The microphone array was mounted atop a noise isolation mast to reduce the effect of local noise generated by Spot's motors and the computer's onboard cooling fans. The system complexity was minimized by assuming deployment in a static environment. A sense-plan-act control procedure was carried out by a finite state machine, as illustrated by Fig. 2.

A. Frontier Exploration Procedure

In this work, Frontier exploration is performed in discrete steps. A global map is built from successive local occupancy grids generated from measurements taken by Spot's surrounding cameras. Occupancy grids span 3 m from all sides of the robot and provide the presence of obstacles at a resolution of 3 cm. The cycle begins by identifying frontiers in the current global map. Frontiers are unoccupied grid cells along the boundary of the known environment. Groupings of frontiers are clustered as a complexity reduction step through adaptive k-means clustering where the number of clusters is calculated by using

$$k = \left\lceil \frac{8m}{n} \right\rceil, \tag{1}$$

where k is the number of clusters used, m is the number of free non-overlapping frontiers, and n is the maximum number of frontiers in an occupancy grid.

A maximum of eight clusters is selected to cover each edge and vertex of the square occupancy grids. Each cluster's centroid is then classified as a Frontier Candidate (FC). FCs are then evaluated by the Frontier Scoring Heuristic (see Section III-B). A trajectory plan is carried out towards the

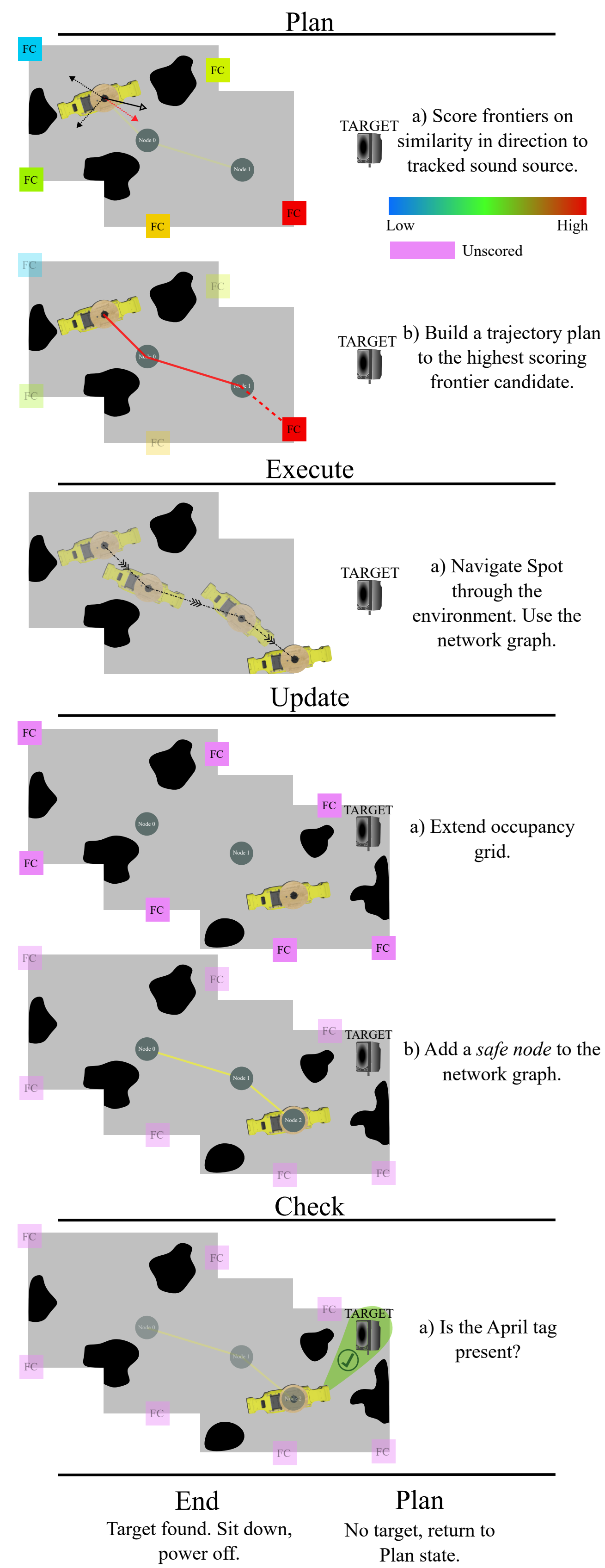

Fig. 2. Illustration of one complete cycle of the system's state machine. Top-down representation of the occupancy grid, network graph, and frontier candidates used to search for the sound source target.

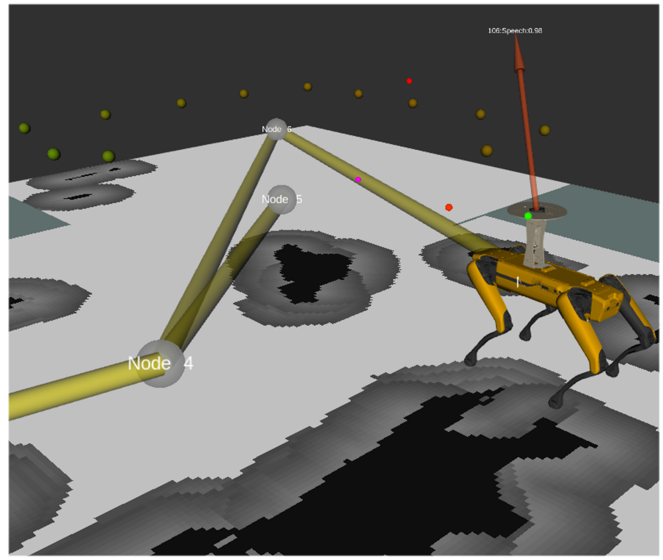

Fig. 3. RVIZ digital twin of Spot and the frontier exploration system. Visualized is the global occupancy map, frontier candidates colour coded by score, the SST unit vector with sound classification, and the network graph of safe nodes.

FC with the highest score. In our mock experiments, upon a successful maneuver, the global occupancy map is extended and the process repeats. Once per cycle, a check state looks for an April tag mounted to the sound source target (Fig. 2 Check). If the April tag is found, the system shuts down and logs run data.

Algorithm 1 Finding a Parent Node for a Frontier Candidate

```
fc ← coordinates of Frontier Candidate
closest ← ∞
node_id ← None
for node in safe_node_graph do
    dist ← euclidean(node, fc)
    if ray_trace_collision_check(node, fc) = 0 then
        approach_possible ← True
    else
        approach_possible ← False
    end if
    if dist < closest and approach_possible then
        closest ← dist
        node_id ← node
    end if
end for
return node_id
```

The global planner employs a network graph of straight edges connecting previously achieved FCs known as *safe nodes* (Fig. 3). Spot maneuvers along the network graph when travelling in the known world. Approaches to FCs are made from the closest *safe node* with an unobstructed straight line path (Alg.1). If an FC approach fails, Spot returns to the last *safe node* and a new trajectory is generated to the FC with the next highest score.

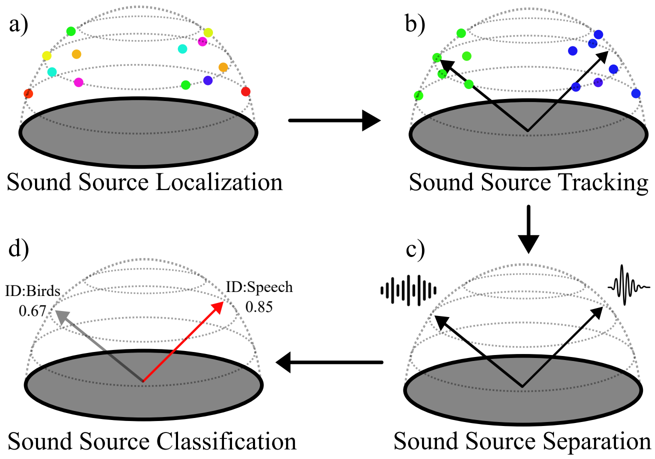

Fig. 4. The four stages of the sound source selection sub-system. Processing raw audio into an estimation of the direction of human voices in the environment.

B. Frontier Scoring Heuristic

FCs are evaluated based on their similarity in direction with the target sound source. A FC is given a score c between -1 and 1 by using

$$c = \vec{f} \cdot \vec{s}, \tag{2}$$

where $\vec{f}$ is a unit vector pointing in the direction of the given FC from the robot's current position and $\vec{s}$ is a unit vector from the same origin pointing towards the sound source target provided by the Sound Source Selection sub-system.

C. Sound Source Selection

The unit vector $\vec{s}$ (Eq. 2) is the result of a four-stage procedure that processes the raw microphone audio into an estimation of the incoming sound direction of human voices present in the environment. The steps involved are:

1) ODAS performs sound source localization by exploiting the microphone array's geometry and the generalized cross-correlation with phase transform method computed for each pair of microphones [6] (Fig. 4a).
2) ODAS applies a Kalman filter to each tracked source. SST detects newly active sources, drops inactive sources, and provides continuous tracking to cope with periods of short silence [6] (Fig. 4b).
3) SST unit vectors are synchronized with their corresponding separated audio track through delay-and-sum beamforming (Fig. 4c).
4) Separated audio tracks are passed through a YAMnet classifier trained on the AudioSet-Youtube corpus [5], assigning a class and confidence rating to each tracked sound source (Fig. 4d)

The sound source with the highest confidence of being human speech is provided to the system as the unit vector $\vec{s}$.

D. Continuous and Discrete Audition

Increased environmental noise reduces the ability of the sound source selection sub-system to identify and accurately track targets at long ranges. Environmental noise was the primary cause of failure in [8], in which only controlled environments were tested. Spot's movements and interactions with different ground materials were the main contributors to local noise. The microphone mast (dubbed "the mushroom") was implemented as a mechanical solution to aid in noise reduction. In addition, we propose two methods of listening:

1) *Continuous:* The microphone is left running. The unit vector $\vec{s}$ is available immediately upon request.
2) *Discrete:* The microphone is selectively turned on when the robot is stationary during each *Plan* state. Noise generated from the robot's movements are effectively eliminated. The unit vector $\vec{s}$ is available once the sound source selection system re-acquires target tracking.

We hypothesized that continuous listening would complete trials quicker because stopping to listen was not required, but discrete listening would complete trials in a shorter distance as it would be less prone to tracking errors from the additional system noise.

IV. FRONTIER EXPLORATION FIELD TRIALS

Experiments were performed on a test field indoors at the Ingenuity Labs Research Institute, at the Kingston Fire and Rescue Training Facility (KFRTF), and at the University of Toronto's Survey Camp (UoTSC) in Minden, Ontario. The indoor test field was an open room where obstacles could be strategically organized into repeatable problems. KFRTF provided two semi-structured urban disaster scenarios, a simulated subterranean concrete access tunnel, and a multi-car traffic accident. Experiments at UoTSC were staged forested environments, carried out in the surrounding unmaintained hiking trials. Locations were chosen to maximize test coverage between controlled laboratory and outdoor experiments.

Ten unique layouts were evaluated with a total of 64 trials being performed (Fig. 5). Test layouts consisted of the Spot robot and a sound source target, separated by an arrangement of obstacles referred to as the *problem*. The sound source target is a public announcement speaker with an attached April tag playing an articulate audio book recording at a measured 91 dBA (observed by the researchers to be at an equal level to a person yelling). Test environments were assumed to be traversable by the Spot platform and contained a single controllable sound source target. Uncontrollable noises were present during outdoor field trials, including noise from other researchers, small animals (birds), high winds, and rustling trees. A minimum of three trials were performed for each audition approach (continuous and discrete), except under conditions where the environment jeopardized the equipment or the acoustics were too poor for either system.

The performance of the system was evaluated by using two metrics: success rate (SR) and success rate weighted by normalized inverse path length (SPL). SR and SPL are widely used in the evaluation of autonomous search agents, described in [1] and applied in [15]. SR was the only quantitative evaluation method extracted from recounts of DR deployments in [10]. The April tag mounted to the

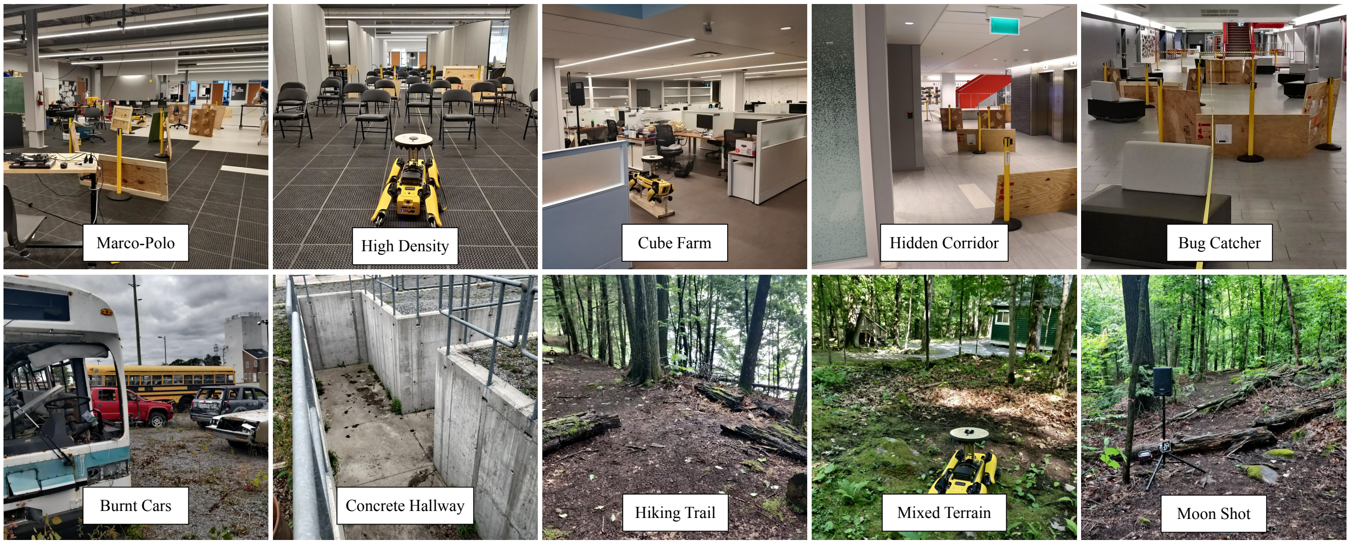

Fig. 5. The ten problems used to evaluate the performance of the system. Urban indoor, urban outdoor, and natural outdoor environments were tested. Each problem was assigned a unique reference name.

sound source target acted as an ObjectGoal. A successful trial was achieved when the system localized the April tag, transitioned to its *end* state, and performed an automatic shutdown. A trial was marked a failure when the system transitioned to its *error* state or an emergency stop was initiated by the researchers.

The comparative analysis of continuous and discrete audition was also analyzed by using Run Time (RT) and overshoot distance (OS). RT is the measured time between system activation by the researchers and the system shutdown down time registered automatically by the system upon the success or failure of a trial. OS is the distance the system traveled beyond the shortest path distance between the starting point and the target, measured by Spot's odometer.

V. EXPERIMENTAL RESULTS

Outcomes of the 64 trials are provided in Table I, with performance of each listening method presented for each of the 10 problems. To our knowledge, this research presents the furthest localization and tracking of human voices for search and rescue robotics at 32.6 m. Previous applications of robotic audition for search and rescue state a maximum range of 15.0 m [12].

A. Overall System Performance

Across all trials, the system achieved a SR of 69 % and a SPL score of 0.42. Indoor and outdoor performance was comparable, with indoor and outdoor testing achieving success rates of 67 % and 70 % respectively and having an identical SPL score of 0.42. The 3 % performance delta in SR and identical SPL score indicates the system can generalize to most problems that are within the capabilities of both the Spot robot and the sound source selection sub-system.

Historical data places the SR of deployed DRs at 50 % (air, land, and sea), dropping to 30 % when considering only unmanned ground vehicles [10, Chapter 2]. Direct comparison between the SR of actual deployments and our simulated search and rescue scenarios does not provide adequate proof of system functionality in search and rescue missions. Our staged worlds, where the system could be evaluated against problems with a intermediate level of repeatability and the expectation of system failure, aligns our testing best with Murphy's definition of *concept experimentation* [10].

Anderson et al. state SPL to be a stringent measurement of navigation performance [1]. Evaluation conducted in unstructured, complex environments with no a priori information should aim for a SPL score of 0.5.

High success rate across all trials shows that directional similarity between FCs and the tracked sound target is a capable heuristic for frontier exploration in short range search and rescue scenarios. An SPL score with a delta of 0.08 from what is considered good navigational performance indicates there are improvements to be made. Field notes remark switching from an ObjectGoal to an AreaGoal as the *Check* state often failed to identify the system had achieved its target, lowering the SPL score by continuing its search. Our trials conclude that sound source tracking as a frontier exploration heuristic in search and rescue is a viable concept for human voice-prioritized search in disaster events.

B. Evaluation of Continuous and Discrete Audition Methods

Continuous and discrete listening methods were evaluated based on trial performances from six of the 10 problems. Circumstances of the trials caused problems *Concrete Hallway*, *Hiking Trail*, *Mixed Terrain*, and *Moon Shot* to be evaluated using only continuous or discrete listening. *Concrete Hallway* and *Moon Shot* environments were beyond the capabilities of the equipment in either listening mode. Dense forest present in the *Moon Shot* problem caused multiple physical failures with the Spot platform. The highly reverberant environment in the *Concrete Hallway* problem

TABLE I

PER PROBLEM SUMMARY OF CONTINUOUS (C.) AND DISCRETE (D.) SUCCESS RATE (SR), SUCCESS RATE PER PATH LENGTH (SPL), RUN TIME (RT) AND OVERSHOOT (OS). HIGHLIGHTING INDICATES THE HIGHER PERFORMING LISTENING METHOD PER PROBLEM FOR EACH PERFORMANCE METRIC.

Problem	C. SR	D. SR	C. SPL	D. SPL	C. RT(s)	D. RT (s)	C. OS (m)	D. OS (m)
Marco-Polo	86%	75%	0.70	0.60	148.1	156.5	9.5	7.9
High Density	67%	50%	0.33	0.31	235.5	226.2	29.7	23.7
Cube Farm	43%	75%	0.25	0.18	90.1	214.4	9.7	44.7
Hidden Corridor	100%	50%	0.52	0.34	284.3	280.9	63.6	16.6
Bug Catcher	100%	75%	0.40	0.52	246.5	210.4	61.1	12.4
Burnt Cars	100%	67%	0.56	0.57	152.7	87.3	15.2	1.1
Comparative 6 Average	78%	64%	0.47	0.41	188.8	200.8	29.9	18.4
Concrete Hallway	33%	N/A	0.23	N/A	114.1	N/A	6.5	N/A
Hiking Traik	N/A	100%	N/A	0.31	N/A	211.8	N/A	53.1
Mixed Terrain	N/A	100%	N/A	0.85	N/A	211.8	N/A	6.8
Moon Shot	0%	N/A	0.00	N/A	N/A	N/A	N/A	N/A
Complete Average	69%		0.42		190.7		20.93	

caused failures in ODAS and the ensuing sound source selection sub-system. Field adjustments to the sound source selection sub-system left continuous listening disabled for the *Hiking Trail* and *Mixed Terrain* trials. Evaluating listening methods through problems that included a minimum of three trials per method presents a clear difference in performance.

Continuous listening had a 14 % higher success rate with an SPL delta of 0.06 over discrete listening. Initial hypotheses are confirmed, continuous listening shows a quicker average run time but a greater distance travelled. The evaluation of the comparative six problems (Table I) identifies continuous listening as the superior approach to sound source selection. Activation of the sound source selection sub-system while the robot is stationary did not improve the systems ability to achieve the target.

VI. CONCLUSION

This system-focused paper introduces the novel application of robotic audition as a heuristic for frontier exploration in search and rescue scenarios. Motivated by the natural human ability to locate people calling for help and the limited capabilities of current teleoperated disaster robots, we utilize a Boston Dynamics Spot robot and a four microphone array to autonomously explore unknown environments using human voice prioritized search. A sense-plan-act control procedure was carried out by a finite state machine. The robot's understanding of its surroundings was progressively expanded by traveling to the boundaries of the explored regions referred to as frontiers. Frontiers were chosen for exploration based on their similarity in direction with a tracked sound source. The sound source was tracked through a four-stage sound source selection sub-system which took in raw audio and provided an estimation of the direction of human voices present in the environment.

The system was proven to generalize well when evaluated against a set of ten distinct problems. Problems included staged urban and forest based search and rescue scenarios. The system achieved a success rate of 69 % and a SPL rating of 0.42, suggesting that sound source tracking is a capable heuristic for frontier exploration at a concept experimentation level of field work for disaster robot applications. Of the two audition methods proposed, continuous listening achieved a 14 % higher success rate and a 6 % higher SPL score. Halting the robot to eliminate noise generated by its movement did not improve the system's ability to achieve the target.

While the system presented in this paper has proven capable in the conducted simulated disaster scenarios, it would require much further development to be considered for deployment in the search for disaster victims. This research stands to demonstrate that sound source tracking tuned for the human voice is a suitable heuristic to bring autonomous search agents used in disaster scenarios closer in capability to the humans they are assisting. Future work should incorporate sound source tracking as a weighted component in a multi-heuristic system, switching search behaviors when a victim is heard nearby. Extending the functionality of this system to track other signals, such as those emitted by firefighter distress beacons, could usher in new possibilities for its use.

ACKNOWLEDGMENT

Special thanks to Thomas Sears for many useful discussions throughout this project, as well as Sophie Villemure and Timothy Byles-Ho for support during field trials.

REFERENCES

[1] Peter Anderson, Angel Chang, Devendra Singh Chaplot, Alexey Dosovitskiy, Saurabh Gupta, Vladlen Koltun, Jana Kosecka, Jitendra Malik, Roozbeh Mottaghi, Manolis Savva, et al. On evaluation of embodied navigation agents. *arXiv preprint arXiv:1807.06757*, 2018.

[2] Ana Batinovic, Tamara Petrovic, Antun Ivanovic, Frano Petric, and Stjepan Bogdan. A multi-resolution frontier-based planner for autonomous 3d exploration. *IEEE Robotics and Automation Letters*, 6(3):4528–4535, 2021.

[3] Amanda Bouman, Muhammad Fadhil Ginting, Nikhilesh Alatur, Matteo Palieri, David D. Fan, Thomas Touma, Torkom Pailevanian, Sung-Kyun Kim, Kyohei Otsu, Joel Burdick, and Ali-akbar Agha-Mohammadi. Autonomous spot: Long-range autonomous exploration of extreme environments with legged locomotion. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2518–2525.

[4] A. Davids. Urban search and rescue robots: from tragedy to technology. *IEEE Intelligent Systems*, 17(2):81–83, 2002.

[5] Jort F. Gemmeke, Daniel P. W. Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R. Channing Moore, Manoj Plakal, and Marvin Ritter. Audio set: An ontology and human-labeled dataset for audio events. In *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, 2017.

[6] François Grondin, Dominic Létourneau, Cédric Godin, Jean-Samuel Lauzon, Jonathan Vincent, Simon Michaud, Samuel Faucher, and François Michaud. Odas: Open embedded audition system. *Frontiers in Robotics and AI*, 9, 2022.

[7] François Grondin and François Michaud. Lightweight and optimized sound source localization and tracking methods for open and closed microphone array configurations. *Robotics and Autonomous Systems*, 113:63–80, 2019.

[8] Ren C. Luo, Chien H. Huang, and Chun Y. Huang. Search and track power charge docking station based on sound source for autonomous mobile robot applications. In *Proceedings of the 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1347–1352.

[9] Robin R. Murphy, Satoshi Tadokoro, Daniele Nardi, Adam Jacoff, Paolo Fiorini, Howie Choset, and Aydan M. Erkmen. Search and rescue robotics. In Bruno Siciliano and Oussama Khatib, editors, *Springer Handbook of Robotics*, pages 1151–1173. Springer.

[10] R.R. Murphy. *Disaster Robotics*. Intelligent Robotics and Autonomous Agents series. MIT Press.

[11] Kazuhiro Nakadai, Makoto Kumon, Hiroshi G. Okuno, Kotaro Hoshiba, Mizuho Wakabayashi, Kai Washizaki, Takahiro Ishiki, Daniel Gabriel, Yoshiaki Bando, Takayuki Morito, Ryosuke Kojima, and Osamu Sugiyama. Development of microphone-array-embedded UAV for search and rescue task. In *Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5985–5990.

[12] Kazuhiro Nakadai and Hiroshi G. Okuno. Robot audition and computational auditory scene analysis. 2(9):2000050.

[13] Kazuhiro Nakadai, Hiroshi G. Okuno, and Takeshi Mizumoto. Development, deployment and applications of robot audition open source software hark. *Journal of Robotics and Mechatronics*, 29(1):16–25, 2017.

[14] Kazuhiro Nakadai, Hiroshi G. Okuno, Hirofumi Nakajima, Yuji Hasegawa, and Hiroshi Tsujino. An open source software system for robot audition hark and its evaluation. In *Humanoids 2008 - 8th IEEE-RAS International Conference on Humanoid Robots*, pages 561–566, 2008.

[15] Jeongeun Park, Taerim Yoon, Jejoon Hong, Youngjae Yu, Matthew Pan, and Sungjoon Choi. SOCRATES: Text-based human search and approach using a robot dog. In *Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, London, England, May 2023. to be published.

[16] B. Yamauchi. A frontier-based approach for autonomous exploration. In *Proceedings 1997 IEEE International Symposium on Computational Intelligence in Robotics and Automation CIRA'97. 'Towards New Computational Principles for Robotics and Automation'*, pages 146–151.