Self-Supervised Learning of Motion-Induced Acoustic Noise Awareness in Social Robots

Submitted: 20th February 2019; accepted: 2nd April 2019

João Andrade, Pedro Santana, Alexandre P. Almeida

DOI: 10.14313/JAMRIS_1-2019/1

Abstract:
With the growing presence of robots in human populated environments, it becomes necessary to render their presence natural, rather than invasive. To do that, robots need to make sure the acoustic noise induced by their motion does not disturb people nearby. In this line, this paper proposes a method that allows the robot to learn how to control the amount of noise it produces, taking into account the environmental context and the robot's mechanical characteristics. Concretely, the robot adapts its motion to a speed that allows it to produce less noise than the environment's background noise and, hence, avoiding to disturb nearby humans. For that, before executing any given task in the environment, the robot learns how much acoustic noise it produces at different speeds in that environment by gathering acoustic information through a microphone. The proposed method was successfully validated on various environments with various background noises. In addition, a PIR sensor was installed on the robot in order to test the robot's ability to trigger the noise-aware speed control procedure when a person enters the sensor's field of view. The use of a such a simple sensor aims at demonstrating the ability of the proposed system to be deployed in minimalistic robots, such as micro unmanned aerial vehicles.

Keywords: *Social Robots, Acoustic Noise, Motion Control, Self-Supervised Learning*

1. Introduction

Robot safe navigation in human-populated environments is one of the most studied topics in the field of robotics since its early days, having reached a point in which self-driving cars became a reality [21]. To be well accepted in environments like offices, households, and factories, robots need to navigate among people in a predictable, non-disturbing way. Socially-aware robot navigation is a relatively new field that aims at exactly solving this problem by including in the robot's navigation system explicit knowledge of human behaviour, including cultural preferences [14, 25].

Social awareness in robots should include geometric aspects of motion planning (e.g., avoid invading personal spaces) but also more subjective aspects related to the acoustic impact robots have over the environment. This is in line with current knowledge about the relevance of adequate noise prevention and mitigation strategies for public health [6]. Hence, depending on the context in which the robot is immersed in, e.g., library versus cafeteria, the robot should be allowed to induce more or less acoustic noise in the environment.

To control the amount of acoustic noise induced in the environment, robots may change their motion accordingly. To that purpose, robots should be provided with a forward model that could predict how much noise will be induced in the environment if a given speed is chosen. With that model the robot should be able to select the speed that induces an acoustic noise level that better trades-off the navigation goal and the comfort of the humans sharing the same environment. This paper proposes a method that allows robots to learn and use these forward models in a way that the robot induces a limited level of acoustic noise trading-off some noise level with the desired speed criteria. Fig. 1 illustrates a typical use-case of the proposed system.

In the proposed system, learning takes place by having the robot performing a set of predefined motor actions to actively induce acoustic noise in the environment. The outcome of these controlled interactions is a set of context-action-sensation tuples that the robot accumulates in an associative memory to learn how to predict its motion-induced noise, given a motor action and an environment context. With the knowledge acquired with this self-supervised active learning strategy, the robot can then select, at each moment, the maximum velocity possible (up-to a reference desired velocity) that induces less acoustic noise than the background's environment acoustic noise. Fig. 2 illustrates the basic principles of operation of the proposed system.

To validate the proposed method, a small-sized ROS-enabled [24] research-oriented wheeled robot, TurtleBot2, has been equipped with a microphone and a simple PIR sensor capable of binary detection of a person in its field of view. The simplicity of the sensory apparatus aims at matching the one that would be available if a small-sized robot would be considered, such as a micro unmanned aerial vehicle. With this approach we intend to demonstrate that the proposed method could be used to such small sized robots, which are expected to populate our environments and possibly organised as swarms (refer to [7] for a survey on swarm robotics). In fact, self-supervised learning in micro air vehicles has been demonstrated in a different context [29].

This article is an extended and improved version of a previously published conference paper [1], providing a more detailed description of the proposed system alongside a deeper analysis of the experimental

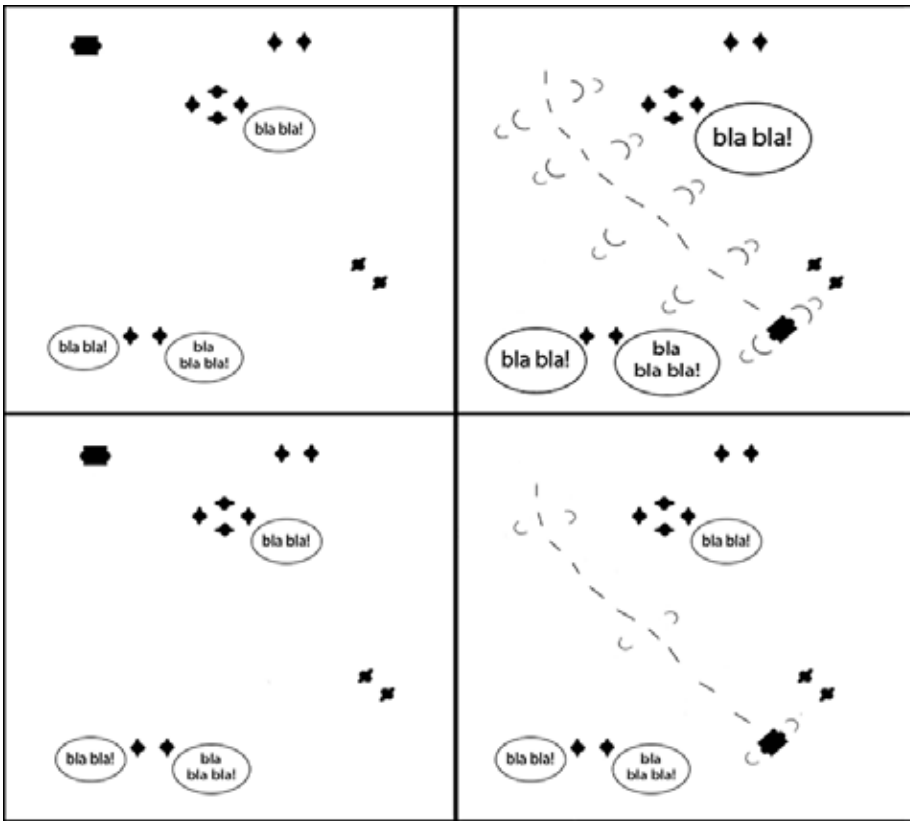

Fig. 1. A cartoon representation of the proposed method's use case. Both top-left and bottom-left images represent the initial state where the robot is idle at an environment where there are some people having conversations. The difference is that at the top images, the robot does not use any noise controller, while in the bottom images, the robot uses the proposed noise controller. At the top-right image, the robot moved near humans while making considerable noise, causing discomfort and forcing people to speak higher to continue the conversations. At the bottom-right image, because the robot is moving slower, it does not make more noise than the people and executes its task while people continue to talk to each other normally. The squared object with two semi circle at the sides is the robot. The humans are represented by circle with a more flatten circle, somewhat similar to a plus signal. The dialogue balloons represent the conversations and the bigger the balloon and the letters, the higher the people are talking. The dotted lines represent the path of the robot and the semi circles nearby represent the noise the robot is producing, where the higher quantity of semi circles, the more noise

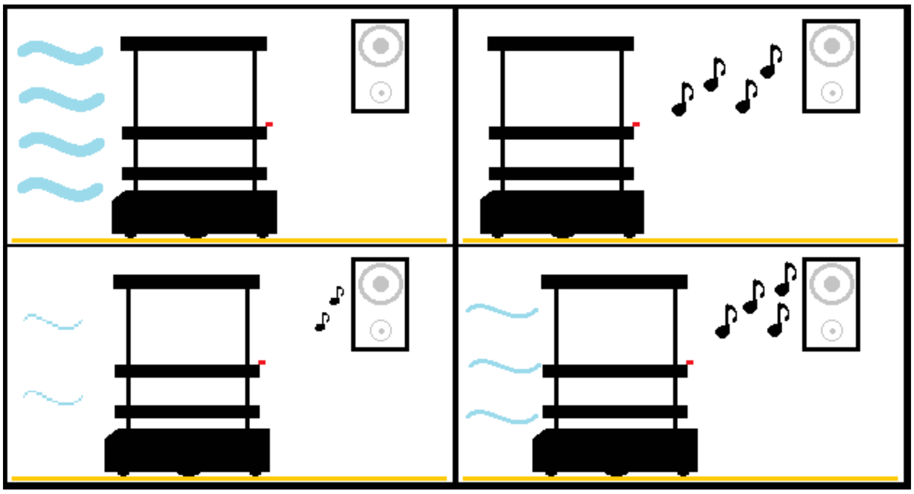

Fig. 2. A cartoon representation of the proposed method's major steps. The red dot at the robot's middle compartment represents the microphone. The blue curved lines represent the robot's speed. The more and ticker the lines, the faster the robot is moving. The musical notes represent the environment's background noise. The more and bigger notes, the higher the volume. At the top-left image, the robot is not using the purposed method, and so, the robot is moving at any speed without consideration of the noise around it. When using the system, the robot needs to be idle and listen to the background noise, as depicted at the top-right image. Then, the lower the background noise (bottom-left image), the slower the robot moves. If there is a high background noise (bottom-right image), the robot is allowed to move faster

results. This paper is organized as follows. Section 2 describes the related work. Section 3 gives an introduction to some theory about sound that inspired this work. Section 4 describes the proposed method and how it can be implemented. Section 5 describes the developed system. In section 6, a set of experimental results are presented. Finally, conclusions and future research avenues are presented in section 7.

2. Related Work

The ability for robots to navigate safely in human-populated environments has been extended in recent years to also encompass human safety, which means that these robots' navigation systems need to be socially-aware. In fact, socially-aware robot navigation has been demonstrated in office environments, houses, and museums [14, 25] and more recently in factories [17, 20].

In addition to ensure the safety of people and goods nearby robots, it is also important to foster comfort in human-robot interactions, as prescribed in the theory of proxemics for human-human interactions [9]. This theory predicts that comfort is a function of the distance between the interacting agents, as well as their relationships, cultures, and intentions. Contemporary socially-aware robot navigation has included these concepts to promote more natural human-robot interactions [16, 25, 27].

Acoustic pollution induced by robots also affects human comfort. A strategy to reduce the acoustic impact robots may have in people-populated environments is to compel these robots to select paths that move closer to sound sources present in the environment [18]. By doing this, the robot masks its acoustic signature with the ones of the sound sources distributed throughout the environment. To handle several acoustic noise sources, acoustic maps can be created and updated by the robot by actively moving in the environment [12, 19]. These maps indicate the location of the several sound sources, which can then be used to hide the robot's acoustic signature. Another application based on the sound captured by a robot's microphones is the detection of obstacles that are outside of its field of vision [13]. This paper contributes to the state of the art by proposing a solution that does not require the explicit mapping of sound sources and, thus, reducing computational complexity and learning time.

The wide variety of environments and robot mechanical structures render difficult designing by hand a set of rules that helps the robot to control its acoustic signature in a way that people do not feed disturbed by its presence. An alternative to the hand crafting of these rules is to allow the robot to learn them in a self-supervised way as a function of the environment and motion speed.

Self-supervised learning has been attracting considerable attention, in particular in safe navigation domain, which requires the robot to autonomously learn classifiers for terrain assessment from images and point clouds [3, 4, 10, 22, 23, 29, 30]. In general, in this previous work the robot is asked to learn what perceptual features better predict a given robot-terrain interaction, provided ground-truth labels produced by an active perception process. For instance, by manipulating an object, the robot is able to obtain ground-truth regarding how traversable that object is [4]. The learned associative mapping can then be used to predict future robot-terrain interactions, given sensory feedback. In this paper we address a similar problem: to learn the acoustic noise induced by the robot in a given environment by engaging in pre-defined motor actions to generate sufficient ground-truth labels for the learning process to take place.

With a strong connection to the ideas of active perception [2, 5], the self-supervised learning concept follows the *affordance* principle studied by Gibson for the animal kingdom [8]. The concept of affordances links the ability of a subject, through its actions, to the features of the environment and, so, to learn an affordance the agent needs to interact with the environment. This idea has been deeply studied in humans [15, 26] and more recently in robotics [11] including for safe navigation purposes [28]. In this paper we address the problem of learning what acoustic noise level is afforded by the environment, given its and the robot's characteristics.

3. Preliminaries on Acoustics

From a physics perspective, sound is a vibration that typically propagates as an audible wave of pressure. This wave propagates through a transmission medium such as a gas, liquid or solid. Our human ears detect changes in sound pressure.

It is well known that the sound level decreases non-linearly as the distance from the sound source increases. Moreover, the characteristics of the environment, e.g., the design of the room (shape, furnishing, surface finishes etc.) influences the extent to which the sound level decreases along with the distance.

Sound pressure level (SPL) or acoustic pressure level is a logarithmic measure of the effective pressure of a sound relative to a reference value. Sound pressure level, denoted L_p and measured in dB, is defined by

$$L_p = 20 \log \frac{p}{p_0} [dB], \quad (1)$$

where p is the root mean square sound pressure and p_0 is the reference sound pressure (normally the lowest threshold of human hearing, 20μPa).

Since microphones have a transfer factor or sensitivity given by some value in mV/Pa, in a particular context or environment, we can relate the sample amplitudes acquired by the microphone to the strength of the acoustic signal (pressure level). This can be represented by the simplified (un-weighted) linear sound pressure level (LSPL) given by

$$LSPL = \frac{1}{N} \sum_{i=1}^{N} \mid x_i - \bar{x} \mid, \quad (2)$$

where N is the number of samples per second, x_i is the sampled amplitude, and $\bar{x}$ is the sample mean value.

4. Proposed Method

4.1. Sensing

Since every environment is unique, the noise induced by the robot is also different when navigating in each of those environments. Therefore, the robot needs to select its speed according to the context in which it is. The set of possible contexts the robot may operate is defined as

$$C = \{c_1, c_2, \ldots\}. \quad (3)$$

The robot's speed also affects the noise it produces. Depending on the robot's characteristics, the robot produces different noise levels. Hence, the robot needs to learn the impact of each speed in each context. However, it is not always possible to test all speeds in all environments. Let us define that the set of speeds that have been tried by the robot in a given context $c \in C$ is

$$S^{[c]} = \{s_1^{[c]}, s_2^{[c]}, \ldots\}. \quad (4)$$

As it will be described below, each speed in each environment is tested multiple times (for robustness purposes) by performing a set of fixed action patterns. As a result of the fixed action patterns noise is produced, whose magnitude is measured with the robot's on-board microphone, resulting in a time-series associated to the context $c \in C$ and speed $s \in S^{[c]}$ in question:

$$X^{[c][s]} = \left\{x^{[c][s]}[0], x^{[c][s]}[1], \ldots x^{[c][s]}[n^{[c][s]}]\right\}. \quad (5)$$

4.2. Learning

Learning occurs by storing in an associative memory the average noise level, $\mu^{[c][s]}$, and a conservative measure of the noise level variation (dispersion), $\sigma_+^{[c][s]}$, *conservative noise level variation* hereafter, observed while performing each assessed speed $s \in S^{[c]}$ in context $c \in C$:

$$M = \left\{\left(\mu^{[c][s]}, \sigma_+^{[c][s]}\right), \forall c \in C, \forall s \in S^{[c]}\right\}, \quad (6)$$

where the average noise level for speed $s \in S^{[c]}$ in context $c \in C$ is given by

$$\mu^{[c][s]} = \frac{1}{n^{[c][s]}} \sum_{i=0}^{n^{[c][s]}} x^{[c][s]}[i]. \quad (7)$$

The conservative noise level variation measure is given by the sum of the standard deviation with the standard error of the mean, allowing to take into account the uncertainty that emerges from the sample size:

$$\sigma_+^{[c][s]} = \sigma^{[c][s]} + \sigma_-^{[c][s]}, \quad (8)$$

where the standard deviation of the noise level for speed $s \in S^{[c]}$ in context $c \in C$ is given by

$$\sigma^{[c][s]} = \sqrt{\frac{\sum_{i=0}^{n^{[c][s]}} \left(x^{[c][s]}[i] - \mu^{[c][s]}\right)^2}{n^{[c][s]} - 1}}, \quad (9)$$

and the standard error of the mean of the noise level for speed $s \in S^{[c]}$ in context $c \in C$ is given by

$$\sigma_-^{[c][s]} = \frac{\sigma^{[c][s]}}{\sqrt{n^{[c][s]}}}. \quad (10)$$

4.3. Memory Recall

The associative memory allows the robot to know how much acoustic noise it induces in the environment at different speeds and contexts. This information is then used by the robot to adapt its speed in order to avoid producing noise whose magnitude is higher than the one of the environment's background noise.

Let us assume the robot needs to perform a given task which requires the robot to move at a given desired speed s_r. Then, the robot needs to determine whether it produces less noise than the environment at speed s_r and, if not, what should be its maximum speed in order to fulfil that condition. To perform this analysis the robot needs to consult its memory.

Let us imagine the robot needs to know the expected conservative noise level variation if travelling at a given speed s_r in a given context c. If that speed has been experienced by the robot, then its memory can be used by a direct recall process. However, if that speed has never been experienced, then the robot needs to linearly interpolate from the two closest speeds stored in memory. Formally, in those cases where $s_r \in S^{[c]}$, the conservative noise level variation is obtained with

$$\sigma_r(c, s_r) = \sigma_+^{[c][s_r]}, \quad (11)$$

whereas in those cases where $s_r \notin S^{[c]}$, the conservative noise level variation is instead obtained with

$$\sigma_r(c, s_r) = \psi\left(s_r, s^-, \sigma^{[c][s^-]}, s^+, \sigma^{[c][s^+]}\right), \quad (12)$$

with:

$$\psi(x, x_0, y_0, x_1, y_1) = \frac{y_0(x_1 - x) + y_1(x - x_0)}{x_1 - x_0}, \quad (13)$$

where the immediately above and below speeds memorised in $M^{[c]}$ for for context $c \in C$ are given by

$$s^{+[c]} = \arg \min_{s \in S^{[c]}, s > s_r} (s - s_r), \quad (14)$$

$$s^{-[c]} = \arg \min_{s \in S^{[c]}, s_r > s} (s_r - s). \quad (15)$$

The robot can also consult its memory to obtain the expected average noise level if travelling at speed s_r. Instead of consulting directly the memory, the robot uses a model learned from the data stored in memory. A set of tests (see below) show that for low speeds a second degree polynomial fits well the data, where as for high speeds a simpler linear model is sufficient. Hence, the expected average noise level is given by,

$$\mu_r(c, s_r) = \begin{cases} ax^2 + bx + c & \text{if } s_r <= s_d \\ dx + e & \text{if } s_r > s_d \end{cases} \quad (16)$$

where $a, \ldots, e$ are parameters learned with regression based on a set of data points corresponding to tuples speed-noise:

$$D^{[c]} = \left\{ \left(s, \mu^{[c][s]} \right), \forall s \in S^{[c]} \right\}. \tag{17}$$

4.4. Motion Control

Algorithm 1) outlines the process used by the proposed system to select which speed is sent to actuators in order to best take into account the robot's induced noise level, its memory, and desired speed.

Data: desired speed, s_r (input)
Data: environment context, c (input)
Data: speed search step, α (input)
Result: Noise aware robot speed control
1 Set robot's speed to 0
2 Store environment noise level for δ seconds in E
3 Compute: $\mu_e = \sum_{e \in E} e/|E|$
4 Initialise x: $x \leftarrow \mu_r(c, s_r) + \sigma_r(c, s_r)$
5 Initialise selected speed: $s \leftarrow s_r$
6 **while** $x > \mu_e \wedge s > \alpha$ **do**
7 | Decrement selected speed: $s \leftarrow s - \alpha$
8 | Update x: $x \leftarrow \mu_r(c, s) + \sigma_r(c, s)$
end
10 Set robot's speed to s

Algorithm 1: Motion controller

The algorithm receives the robot's desired speed, s_r, if noise level was not a concern. This speed is often task-oriented. The algorithm also assumes that the environment context, c, is known, for instance using vision (not the focus of this article). The output of this algorithm is, if possible, the highest speed, up to s_r, the robot can move that does not produce more noise than the environment.

First, the robot is asked to stop (Step 1). This way, the robot's induced noise does not interfere with the next step. With the microphone, the robot gathers the environment's noise levels, E, for a determined number of seconds, δ (Step 2). These noise levels are then used to calculate an average background noise level, μ_e (Step 3). Then, the robot predicts, in a conservative manner, the expected noise produced by the robot at speed s (Step 4).

The selected speed, s, is initially set to the desired speed s_r (Step 5), because the desired speed must be the highest selected speed possible. Then, a small cycle (Steps 6-9) needs to be performed to find the best speed. While the predicted robot's noise is higher than the environment's background noise and the selected speed is higher than a given speed search step α (Step 6), the selected speed is decremented (Step 7) and the predicted robot's noise with that speed is computed (Step 8).

When the robot finds a speed which is expected to produce less noise than the environment's, the algorithm ends and the robot actuator speed is set to the selected speed s (Step 10). If the selected speed gets lower than the speed search step, then it is assumed

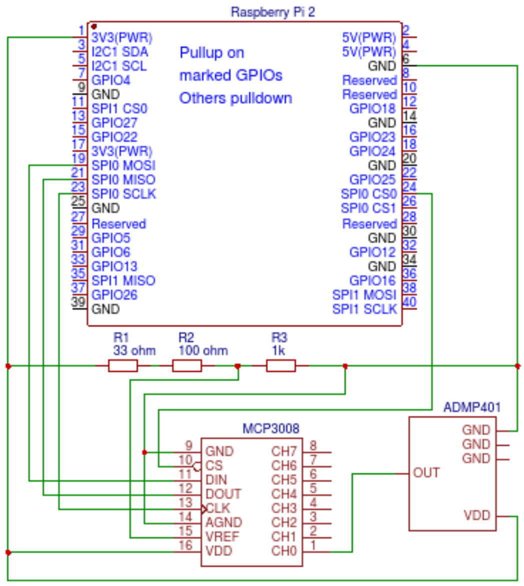

Fig. 3. Diagram showing the connections between the microphone and the Raspberry Pi 2. It is possible to see that the 'Vref' (reference voltage) is different from the "VDD" on the ADC in order to amplify a bit the signal

the robot cannot produce less noise than the environment and the robot's speed is set to the selected speed. Hence, the speed search step α should be set to the minimum speed possible the robot can perform the task at hand.

5. Experimental Setup

The proposed method has been devised to allow complex and minimalist robots to adjust their speed in order to control their induced noise when navigating nearby people (e.g., a micro aerial vehicle flying in a office environment). However, the system could be easily adapted in order to perform the opposite operation, that is, render the robot salient in the acoustic landscape. This could be interesting for tasks in which would be important to attract people's attention to the robot.

This section presents the instantiation of the proposed method to a small-sized ground robot, a TurtleBot 2.0, equipped with a microphone and a Raspberry pi 2 Model B for data acquisition and transmission to the robot's main processing unit. Fig. 3 shows a digram with the connections between the microphone and the raspberry pi, whereas Fig. 4 depicts the robot used.

5.1. Microphone position

The robot has three compartments in which the microphone can be placed (top compartment, middle compartment and bottom compartment), as depicted in Fig. 5. In addition to selecting the most appropriate compartment for microphone placement, it is necessary to verify whether it should be placed at the front or at the back of the robot, making a total of six different possible positions. In order to determine the best position, an experiment was performed. In all tested positions, the microphone is pointing downwards so

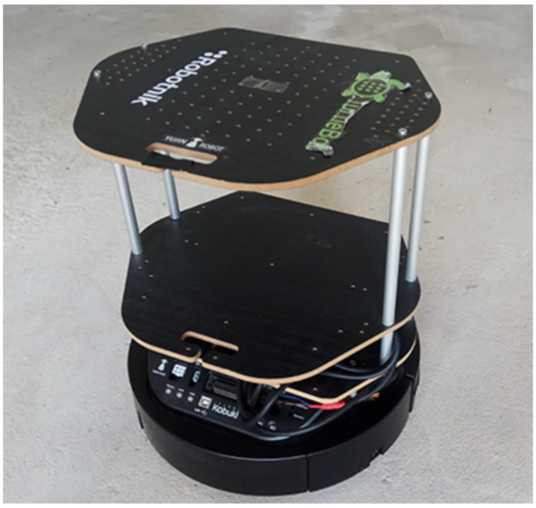

Fig. 4. Turtlebot 2.0. The robot used in this work

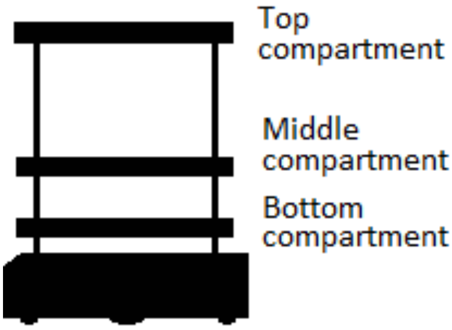

Fig. 5. Robot with its top, middle, and bottom compartments

as to be highly influenced by the wheels and motors acoustic noise. Fig. 6 plots the acoustic noise level recorded by the robot's microphone when placed in each of the six positions while travelling at 0.2 m/s.

Fig. 6 shows that there was no major difference between having the microphone at the back or at the front of the robot in any compartment, although it has a little better performance at the back. Regarding the different compartments, the top compartment provides the worst performance. It is possible to differentiate the robot's noise, but not as well as in the others compartments. Both the middle and bottom compartment are good to distinguish the robot's noise, with the bottom having an advantage because it is more perceptible.

As the robot can potentially move at higher speeds than the tested 0.2 m/s, for example in cases where there is more noise in the environment, a second experiment with the robot moving at 0.5 m/s on a concrete floor was performed. Fig. 7 shows the results of that experiment.

It is noticeable in Fig. 7 how much more the noise is saturated when the microphone is at the bottom compartment. Saturation occurs due to closeness to the motors and due to the high mechanical impact induced by the rough terrain on the robot's structure. For this reason, the back of the middle compartment was selected as the most appropriate to place the microphone. The remainder of the article assumes the microphone to be located in this selected position.

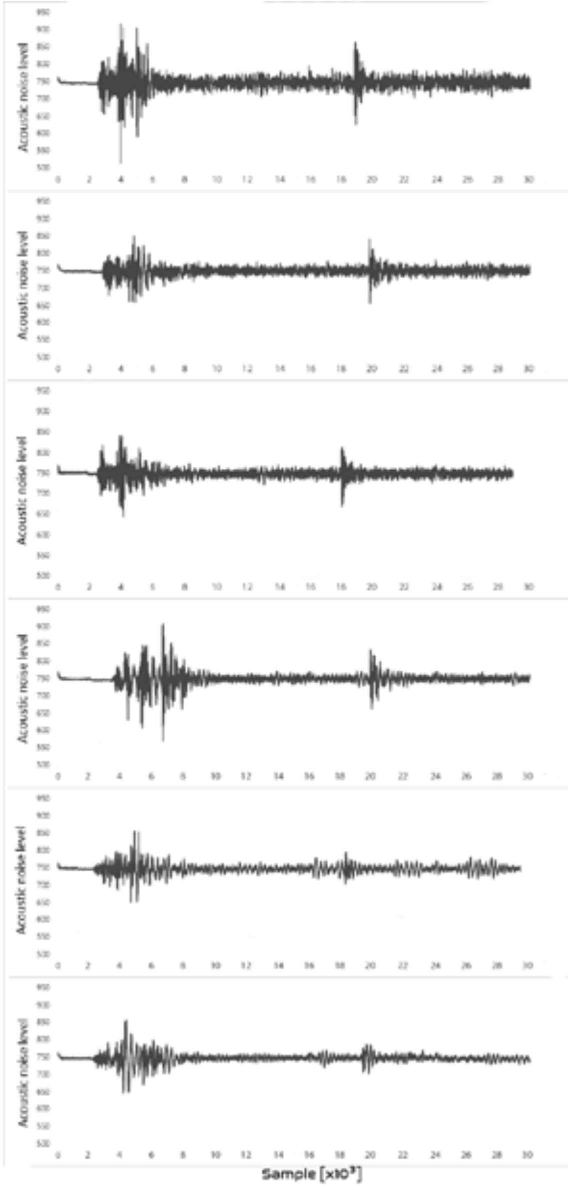

Fig. 6. Acoustic noise level induced by the robot while moving at 0.2 m/s on top of ceramic tiles, with the microphone on different compartments and located at the front and back of the robot. It is possible to check the environment's noise right at the beginning. The spike that follows represent the robot's motion onset. Then the noise stabilizes because the robot is constantly moving at the target speed. The spike that occurs at the middle represent the moment when the robot passes from one ceramic tile to another, which causes some mechanical impact that results in a higher acoustic noise. The following order represents the images sequentially from top to bottom: 1 - bottom compartment, back position; 2 - bottom compartment, front position; 3 - middle compartment, back position; 4 - middle compartment, front position; 5 - higher compartment, back position; 6 - higher compartment, front position. The vertical axis represents the acoustic noise level, and the horizontal axis represents the different samples from the microphone

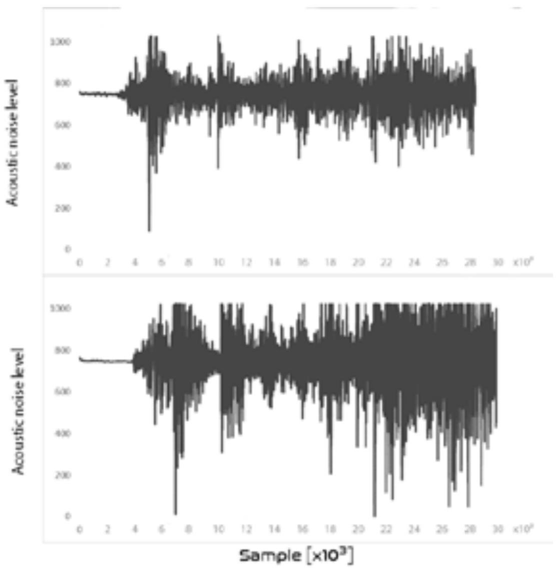

Fig. 7. Acoustic noise level induced by the robot while moving at 0.5 m/s on top of rough concrete floor, with the microphone on different compartments located at the back of the robot. Like on fig. 4, the vertical axis represents the acoustic noise level, and the horizontal axis represents the different samples from the microphone. The top image plots the samples acquired with the microphone on the middle compartment and the bottom image plots the samples acquired with the microphone on the bottom compartment

5.2. Learning phase

With the microphone position established, it is possible to gather the robot's noise. To test if the system works in various scenarios, four different contexts (environments) were selected. Since the robot moves with wheels, the major difference between the contexts is the floor (tiles, cement, carpet and wood). The set of tested contexts is:

$$C = \{tiles, wood, cement, carpet\}. \tag{18}$$

For a robot moving alongside humans, it cannot move too fast, as that can be uncomfortable for them. For that reason, the set of speeds ranges from 0.0 m/s to 0.8 m/s in 0.1 m/s steps. Gathering data when the robot is stopped (0.0 m/s) can be important in situations where the robot's noise is not as easily distinguishable from the environment's noise, and so, the robot can have a reference of the environment's noise. The set of tested speeds in each environment $c \in C$ is:

$$S^{[c]} = \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}. \tag{19}$$

The robot moves at each speed for each context for about 2 seconds. This is repeated 15 times as to have a good amount of samples. For each context, the conditions are the same. The robot always starts at the same place, to make each repetition as similar as possible, and it needed to be as much silent as possible in the environment. Every time there is some kind of noise that is not from the robot's motion, that run is discarded and a new one is performed until all 15 similar repetitions are completed.

The microphone data gathered for each context and speed compose the set of noise levels ($X^{[c][s]}$), which allows to store the set of tuples (M) that contain the noise level average and the conservative noise level variation and understand the regression equations for each context ($\mu_r(c, s_r)$).

5.3. Motion controller

Placing the microphone at a good place and learning the robot's noise are two necessary steps to make the robot make less noise than the environment. With those two steps done, it is possible to develop the motion controller, which is what the robot uses to know what speed should use.

The way this motion controller works is based on the algorithm presented in the previous section. Before starting, the robot must be idle, as to not pollute the microphone data with its noise and require an extra element on the algorithm to filter its noise. The selected speed (s) is first initialized as the desired speed (s_r) and the search step (α) is set. The robot starts by gathering the environment's noise levels (δ) for 1 second with the microphone, which is then calculated the average value (μ_e) and considered the maximum acoustic noise the robot can produce. After that second, a small cycle is performed. With the help of the data gathered during the learning phase, it is estimated the average noise level (μ_r) and the conservative noise level variation (σ_r). If the environment noise is lower than the average noise plus the conservative noise level variation ($\mu_e < (\mu_r + \sigma_r)$), it is decremented 0.1 to s and the cycle is repeated until either $\mu_e > (\mu_r + \sigma_r)$, where the best velocity is found, or S is lower than α, where the robot is moving too slow to be able to execute a task efficiently or in a reasonable amount of time and it assumes it cannot hide its own acoustic noise.

6. Experimental Results

6.1. Learning Phase

To better understand the different contexts, Fig. 9 shows the average noise values and expected noise level variation that the robot made for each speed-context pair. The figure shows that, because the concrete floor is the hardest floor type, the robot produces more acoustic noise than on the other contexts. Predictably, the carpet floor, the softer floor type, makes the robot produce lesser acoustic noise than the other contexts. As it is possible to see, up until 0.4 m/s there is a higher variation between the noise values than the higher velocities, so it makes sense to have two different regression equations, as described in Equation 16. One of the equations represent the velocities between 0 m/s and 0.4 m/s, whereas the other the velocities between 0.4 m/s and 0.8 m/s.

Fig. 8. Different contexts tested. Top left: Carpet floor; Top right: Tiles floor; Bottom left: Cement floor; Bottom right: Wood floor

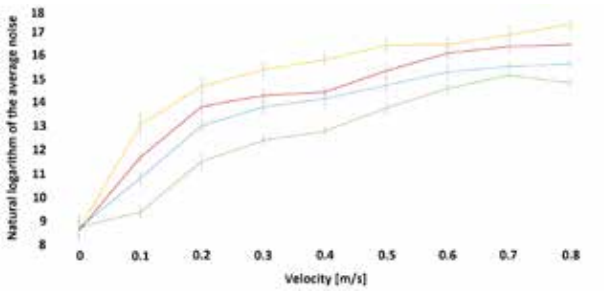

Fig. 9. Learning phase results showing the robot's average noise level and conservative noise level variation for the different contexts. Contexts from top to bottom: cement, tiles, wood and carpet. The horizontal axis represents the speed of the robot in m/s, and the vertical axis represents the average noise of the robot. The vertical lines at each velocity represents the conservative noise level variation

6.2. Motion Controller

To validate the motion controller, a simple yet effective experiment was performed. The robot was placed in all four contexts (tiles, cement, wood and carpet) idle. Then, a set of sound clips started playing. Three different sound clips with two intensity levels (high volume and low volume) were used. Since this is a controlled experiment, the sounds were played by speakers placed in the environment: (a) a sound clip from a crowded area, to simulate an environment where there are people inducing some background noise; (b) a vacuum cleaner, to simulate an environment where there is a constant background noise, like an air ventilation system; and (3) a jazz song to simulate an environment where the volume from the background noise is dynamic, meaning it oscillates its volume based on the song's characteristics. While each sound clips were playing, the motion controller's algorithm was executed to find the ideal velocity s. The speed search step α was set to 0.05 and the desired speed s_r, to force the algorithm to search for the best velocity, was set to a high value of 2.0 m/s. Table 1 shows the results obtained from that experience.

The table show that the predicted acoustic noise the robot should produce is not higher than the environment's background noise with similar values, which means that people nearby should not be disturbed by the robot's noise. Notice that the noise is measured by the microphone on the robot, so the farther away from the robot a person may be, the less impactful is the robot to that person.

6.3. Testing with a People Detector

To further test the system, a second experiment was conducted. This time, the robot was equipped with a PIR sensor to detect the presence of a person and it was performed on a tiled floor. The sound clips used in this experiment are the same as in the previous experiment (crowd sound, vacuum cleaner and a jazz song), except this time, there were no volume variations and all sound clips were producing similar noise levels. The PIR installed is a Motion Sensor Mo-

Tab. 1. Results from the motion controller implementation. The headers are, from left to right: Ambient noise, floor type, environment back noise, predicted robot induced noise, noise difference and noise ratio

Ambient sound	Floor type	Env. back noise (δ)	Pred. rob. ind. noise (μ_r)	Noise difference ($\mu_r - \delta$)	noise ration [%] ($1 - \mu_r/\sigma$)
Jazz High volume	concrete	16.61	16.59	-0.02	0.10%
	tiles	16.63	16.62	-0.01	0.03%
	wood	16.53	16.51	-0.02	0.15%
	carpet	16.03	16.01	-0.02	0.10%
Jazz Low volume	concrete	14.24	14.20	-0.04	0.30%
	tiles	14.28	14.27	-0.01	0.05%
	wood	13.81	13.76	-0.05	0.34%
	carpet	12.81	12.79	-0.02	0.14%
Crowd High volume	concrete	16.61	16.59	-0.02	0.11%
	tiles	16.40	16.38	-0.02	0.14%
	wood	16.04	16.02	-0.02	0.12%
	carpet	15.94	15.93	-0.01	0.08%
Crowd Low volume	concrete	15.64	15.59	-0.05	0.35%
	tiles	13.91	13.84	-0.07	0.51%
	wood	12.99	12.95	-0.04	0.28%
	carpet	11.80	11.79	-0.01	0.09%
Vacuum cleaner	concrete	15.29	15.25	-0.04	0.27%
	tiles	16.15	16.13	-0.02	0.13%
	wood	15.51	15.50	-0.01	0.04%
	carpet	14.75	14.74	-0.01	0.07%

dule IM120628009, which has a range of 7 m and a field of view of 110 degrees, and was installed on the front of the robot's top compartment.

The robot starts by moving forward at a desired speed s_r=0.5 m/s until the PIR sensor detects a person. When a person passes by the PIR field of view, the robot stops and performs the algorithm of the motion controller to find the ideal speed s and starts moving at that that speed. Since this is a controlled experience, the person appears always at approximately 1 m in front of the robot. This test is useful to understand if the robot is capable of performing a task with this motion controller, where the objective is the robot to stop when a person is nearby and adapt its velocity as to not cause discomfort to that person. For example, an autonomous vacuum cleaner could clean an house room slower when a person is in the same division. Table 2 shows the result obtained from this experience.

Similar to the previous experiment, the robot does not produce more acoustic noise than the background environment's noise, with the average difference between the robot's noise and the background's noise being 1.24%, and shows that the motion controller can be integrated into a more complex system to perform different types of tasks.

7. Conclusion

Regardless of the robots activity, the acoustic noise it induces in the environment can be uncomfortable or annoying to people that might be in the same environment. Therefore, it is necessary to limit the amount of acoustic noise produced by the robot so it becomes unobtrusive.

By setting a microphone on a robot, it is possible to learn the amount of acoustic noise any robot makes while moving at any speed in any context. In this work, a system was developed to enable different kinds of robots (big, small, aerial or grounded) to adapt their motion when around humans by moving at a speed that will not produce more acoustic noise than the already present in the environment's background, and consequentially, not cause discomfort to people because of the robot's noise.

To that purpose, the robot needs to have a notion of how much acoustic noise it produces. This is accomplished by having a learning phase in the different contexts where the robot is expected to perform its tasks. This learning phase creates a relation between a tuple (average and conservative variation of the noise level) and the different speeds that the robot may use. This is the only necessary task needed before being able to use the developed system.

The developed system is a motion controller that, at any moment, allows the robot to adapt its speed to a value that does not produce more acoustic noise than the already existing in the environment. This allows for a more acceptance of autonomous systems in our society, because by not disturbing nearby people with its noise, people can perform whatever they may be doing regardless of robots nearby.

Although the results suggest that the proposed method works in a set of disparate contexts, there are some contexts where performing a task without causing some discomfort to humans nearby is almost impossible. Quiet places, like in a library, where there is not much background noise, the robot may not be able

Tab. 2. Results from the second experiment, where the robot stops performing a task with the presence of a person and adapts its velocity to not disturb the person. The headers are, from left to right: Ambient sound, predicted robot induced noise, environment back noise, robot selected speed, noise difference and noise ratio

ambient sound	pred. rob. ind. noise (μ_r)	env. back noise (δ)	robot selected speed (s)	noise difference ($\mu_r - \delta$)	noise ratio[%] ($1 - \mu_r/\sigma$)
Vac. cleaner	0.395	14.89	14.57	-0.32	2.12%
Crowd	0.405	14.96	14.94	-0.02	0.13%
Jazz	0.395	14.79	14.57	-0.22	1.49%

to execute its tasks at a reasonable speed. In those situations, the robot or the user controlling it will have that knowledge and may move at the lowest speed possible or postpone the completion of the task until there are no people nearby.

To test the proposed method, a couple of experiments were performed. The first experience involved four different contexts (cement floor, tiled floor, wooden floor and carpet floor) and three different sound clips: Jazz song, crowd noise and a vacuum cleaner. Each sound clip had a different purpose. The jazz song simulated an environment where there are variation in the background's noise volume, the crowd noise simulated an environment with multiple people nearby and the vacuum cleaner simulated an environment where there is a constant background noise. The robot was placed at each context and executed the motion controller to find the highest speed it could move to not make more acoustic noise than the background noise. The experimental results showed that the proposed method properly handled the various situations by selecting speeds that would allow the robot to not produce more noise than the environment. It is worth noting that the noise values are from the robot's perspective, meaning that a person hears the robot's noise more or less depending on the distance it has from the robot.

To further validate this method, a second experiment was performed. The robot, equipped with a PIR sensor, was placed in only one context and the objective was to test the ability of the robot to select a proper speed as soon as someone appeared in the PIR's field of view. The obtained results show that the robot was able to not move at a speed that would make it produce more noise than the environment's background noise.

Despite the overall positive results, the proposed method presents some limitations to be handled in future work. For example, the experimental results were obtained with the robot performing simple forward motions. Future work needs to address a more diverse set of motion primitives. Although tested in different contexts, all were indoors with flat terrains. Future work should assess the system in a wider range of environments (e.g., rough outdoor environments). It would also be valuable to assess the system in robots with different morphologies (e.g., small unmanned aerial vehicles).

AUTHORS

João Andrade – Instituto de Telecomunicações and ISCTE-Instituto Universitário de Lisboa, Lisboa, PORTUGAL, e-mail: joaopcandrade@gmail.com.

Pedro Santana* – Instituto de Telecomunicações and ISCTE-Instituto Universitário de Lisboa, Lisboa, PORTUGAL, e-mail: pedro.santana@iscte-iul.pt.

Alexandre P. Almeida – Instituto de Telecomunicações and ISCTE-Instituto Universitário de Lisboa, Lisboa, PORTUGAL, e-mail: alexandre.almeida@iscte-iul.pt.

*Corresponding author

REFERENCES

[1] J. Andrade, P. Santana, and A. Almeida, "Motion-induced acoustic noise awareness for socially-aware robot navigation". In: *Proceedings of the IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*, 2018, 24–29, 10.1109/ICARSC.2018.8374155.

[2] R. Bajcsy, "Active perception", *Proceedings of the IEEE*, vol. 76, no. 8, 1988, 966–1005, 10.1109/5.5968.

[3] M. Bajracharya, A. Howard, L. H. Matthies, B. Tang, and M. Turmon, "Autonomous off-road navigation with end-to-end learning for the LAGR program", *Journal of Field Robotics*, vol. 26, no. 1, 2009, 3–25, 10.1002/rob.20269.

[4] J. Baleia, P. Santana, and J. Barata, "On exploiting haptic cues for self-supervised learning of depth-based robot navigation affordances", *Journal of Intelligent & Robotic Systems*, vol. 80, no. 3-4, 2015, 455–474, 10.1007/s10846-015-0184-4.

[5] D. H. Ballard, "Animate vision", *Artificial intelligence*, vol. 48, no. 1, 1991, 57–86, 10.1016/0004-3702(91)90080-4.

[6] M. Basner, W. Babisch, A. Davis, M. Brink, C. Clark, S. Janssen, and S. Stansfeld, "Auditory and non-auditory effects of noise on health", *The Lancet*, vol. 383, no. 9925, 2014, 1325–1332, 10.1016/S0140-6736(13)61613-X.

[7] M. Brambilla, E. Ferrante, M. Birattari, and M. Dorigo, "Swarm robotics: a review from the

swarm engineering perspective", *Swarm Intelligence*, vol. 7, no. 1, 2013, 1–41, 10.1007/s11721-012-0075-2.

[8] J. Gibson, "The concept of affordances", *Perceiving, acting, and knowing*, 1977, 67–82.

[9] E. Hall, *The Hidden Dimension : man's use of space in public and in private*, Publisher is empty!, 1969, 217.

[10] H. Heidarsson and G. Sukhatme, "Obstacle detection from overhead imagery using self-supervised learning for autonomous surface vehicles". In: *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2011, 3160–3165, 10.1109/IROS.2011.6094610.

[11] L. Jamone, E. Ugur, A. Cangelosi, L. Fadiga, A. Bernardino, J. Piater, and J. Santos-Victor, "Affordances in psychology, neuroscience and robotics: a survey", *IEEE Transactions on Cognitive and Developmental Systems*, 2016, 10.1109/TCDS.2016.2594134.

[12] N. Kallakuri, J. Even, Y. Morales, C. Ishi, and N. Hagita, "Probabilistic approach for building auditory maps with a mobile microphone array". In: *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, 2013, 2270–2275, 10.1109/ICRA.2013.6630884.

[13] N. Kallakuri, J. Even, Y. Morales, C. Ishi, and N. Hagita, "Using sound reflections to detect moving entities out of the field of view". In: *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, 2013, 5201–5206, 10.1109/IROS.2013.6697108.

[14] T. Kruse, A. K. Pandey, R. Alami, and A. Kirsch, "Human-aware robot navigation: A survey", *Robotics and Autonomous Systems*, vol. 61, no. 12, 2013, 1726–1743, 10.1016/j.robot.2013.05.007.

[15] S. Lacey, J. Hall, and K. Sathian, "Are surface properties integrated into visuohaptic object representations?", *European Journal of Neuroscience*, vol. 31, no. 10, 2010, 1882–1888, 10.1111/j.1460-9568.2010.07204.x.

[16] M. Luber, L. Spinello, J. Silva, and K. O. Arras, "Socially-aware robot navigation: A learning approach". In: *2012 IEEE/RSJ international conference on Intelligent robots and systems (IROS)*, 2012, 902–907, 10.1109/IROS.2012.6385716.

[17] F. Marques, D. Gonçalves, J. Barata, and P. Santana, "Human-aware navigation for autonomous mobile robots for intra-factory logistics". In: *Proceedings of the 6th International Workshop on Symbiotic Interaction*, 2017, 10.1007/978-3-319-91593-7_9.

[18] E. Martinson, "Hiding the acoustic signature of a mobile robot". In: *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, 2007, 985–990, 10.1109/IROS.2007.4399264.

[19] E. Martinson and A. Schultz, "Auditory evidence grids". In: *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, 2006, 1139–1144, 10.1109/IROS.2006.281843.

[20] G. Michalos, S. Makris, P. Tsarouchi, T. Guasch, D. Kontovrakis, and G. Chryssolouris, "Design considerations for safe human-robot collaborative workplaces", *Procedia CIrP*, vol. 37, 2015, 248–253, 10.1016/j.procir.2015.08.014.

[21] B. Paden, M. Čáp, S. Z. Yong, D. Yershov, and E. Frazzoli, "A survey of motion planning and control techniques for self-driving urban vehicles", *IEEE Transactions on Intelligent Vehicles*, vol. 1, no. 1, 2016, 33–55, 10.1109/TIV.2016.2578706.

[22] E. Pinto, F. Marques, R. Mendonça, A. Lourenço, P. Santana, and J. Barata, "An autonomous surface-aerial marsupial robotic team for riverine environmental monitoring: Benefiting from coordinated aerial, underwater, and surface level perception". In: *Robotics and Biomimetics (ROBIO), 2014 IEEE international conference on*, 2014, 443–450.

[23] R. Pombeiro, R. Mendonça, P. Rodrigues, F. Marques, A. Lourenço, E. Pinto, P. Santana, and J. Barata, "Water detection from downwash-induced optical flow for a multirotor uav". In: *OCEANS'15 MTS/IEEE Washington*, 2015, 1–6, 10.23919/OCEANS.2015.7404458.

[24] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "Ros: an open-source robot operating system". In: *ICRA workshop on open source software*, vol. 3, no. 3.2, 2009, 5.

[25] J. Rios-Martinez, A. Spalanzani, and C. Laugier, "From proxemics theory to socially-aware navigation: A survey", *International Journal of Social Robotics*, vol. 7, no. 2, 2015, 137–153, 10.1007/s12369-014-0251-1.

[26] J. Schwenkler, "Do things look the way they feel?", *Analysis*, vol. 73, no. 1, 2013, 86–96.

[27] L. Takayama and C. Pantofaru, "Influences on proxemic behaviors in human-robot interaction". In: *IEEE/RSJ international conference on Intelligent robots and systems, 2009. IROS*, 2009, 5495–5502, 10.1109/IROS.2009.5354145.

[28] E. Uğur and E. Şahin, "Traversability: A case study for learning and perceiving affordances in robots", *Adaptive Behavior*, vol. 18, no. 3-4, 2010, 258–284, 10.1177/1059712310370625.

[29] K. van Hecke, G. de Croon, L. van der Maaten, D. Hennes, and D. Izzo, "Persistent self-supervised learning: From stereo to monocular vision for obstacle avoidance", *International Journal of Micro Air Vehicles*, vol. 10, no. 2, 2018, 186–206, 10.1177/1756829318756355.

[30] K. M. Wurm, H. Kretzschmar, R. Kümmerle, C. Stachniss, and W. Burgard, "Identifying

vegetation from laser data in structured outdoor environments", *Robotics and Autonomous Systems*, vol. 62, no. 5, 2012, 675–684, 10.1016/j.robot.2012.10.003.