High-Resolution Ultrasound Sensing for Robotics Using Dense Microphone Arrays

THOMAS VERELLEN, ROBIN KERSTENS, AND JAN STECKEL, (Member, IEEE)

FTI-CoSys Lab, University of Antwerp, 2000 Antwerp, Belgium
Flanders Make Strategic Research Centre, 3920 Lommel, Belgium

Corresponding author: Jan Steckel (jan.steckel@uantwerpen.be)

ABSTRACT State-of-the-art autonomous vehicles use all kinds of sensors based on light, such as a camera or LIDAR (Laser Imaging Detection And Ranging). These sensors tend to fail when exposed to airborne particles. Ultrasonic sensors have the ability to work in these environments since they have longer wavelengths and are based on acoustics, making them able to pass through the mentioned distortions. However, they have a lower angular resolution compared to their optical counterparts. In this article a 3D in-air sonar sensor is simulated, consisting of a Uniform Rectangular Array similar to the newly developed micro Real Time Imaging Sonar (μRTIS) by CoSys-Lab. Different direction of arrival techniques will be compared for an 8 by 8 uniform rectangular microphone array in a simulation environment to investigate the influence of different parameters in a completely controlled environment. We will investigate the influence of the signal-to-noise ratio and number of snapshots to the angular and spatial resolution in the direction parallel and perpendicular to the direction of the emitted signal, respectively called the angular and range resolution. We will compare these results with real-life imaging results of the μRTIS. The results presented in this article show that, despite the fact that in-air sonar applications are limited to only one snapshot, more advanced algorithms than Delay-And-Sum beamforming are viable options, which is confirmed with the real-life data captured by the μRTIS.

INDEX TERMS Acoustic signal processing, array signal processing, beamforming, microphone arrays, sonar.

I. INTRODUCTION

Autonomous vehicles mostly use optical techniques to perceive their environment. However, these sensors tend to fail when exposed to airborne particles [1], [2]. Therefore, it is of interest to further complement the data gathered with optical techniques with information gathered from ultrasonic sensors. These sensors use acoustical waves with relatively long wavelengths which pass through the propagation medium's distortions. However, there are downsides to using these ultrasonic sensors. Sound waves travel slower in air compared to light (343 m/s compared to 299.8×10^6 m/s), as a result, taking one measurement over a range of six meters takes 35 ms. This makes gathering multiple snapshots impossible since it is important that the environment stays completely stationary, which cannot be guaranteed for mobile applications. Additionally, because of the longer wavelengths and acoustics that are used, the angular resolution cannot compete

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Liu.

with the accuracy of alternatives such as LIDAR or a regular camera [3].

The embedded Real Time Imaging Sonar (eRTIS) is a high-accuracy 3D sonar sensors based on a 32 pseudo-random microphone-array which use Delay-And-Sum (DAS) beamforming to solve the Direction Of Arrival (DOA) problem [4]–[6]. Recently, a smaller version of the eRTIS was developed to satisfy the need for 3D in-air sonar sensors. This new version features a smaller footprint for all kind of robotic applications in need for cost-effective 3D environmental perception which do not rely on optics. This smaller version is named the micro Real Time Imaging Sonar (μRTIS) [7]. The μRTIS features 30 microphones placed in Uniform Rectangular Array (URA) of 5 by 6 microphones in a footprint of 5.7 cm by 4.6 cm.

The original eRTIS used DAS which is known for its ease of use, low computational need, and low angular resolution. To improve the angular resolution, the μRTIS uses MUltiple SIgnal Classification (MUSIC) algorithm [8]. MUSIC requires more computational effort than DAS but

has the prospect to improve the angular resolution of the 3D sonar sensor. However, MUSIC fails to correctly resolve the different sources in coherent-source spaces [9], [10]. This is because the decomposition of the source correlation matrix used by MUSIC requires this matrix to be of full rank. In the presence of K uncorrelated sources, the K highest eigenvalues will correspond with the signal subspace. The eigenvectors corresponding to the remaining eigenvalues make up the noise subspace [11]. In a coherent subspace, the eigenvalues of perfectly coherent signals will be zero and MUSIC will fail to correctly resolve these sources. Since the reflected signals originating from the ultrasonic sensor are all coherent, we will preprocess them by means of spatial smoothing. Spatial smoothing averages the source correlation matrix by dividing the microphones in several identical overlapping arrays, which restores its rank [12]–[14].

The aim of this article is to introduce the μRTIS, the newly developed 3D in-air sensor from CoSys-Lab. In this article we analyze its performance with respect to various beamforming techniques such as DAS, Capon, and MUSIC. We apply these techniques to a simulated signal of an in-air sonar sensor in a completely controlled environment. This simulation environment enables us to more accurately research the influence of different parameters which is not possible in real-life. The results of these simulations can be used in future work to further improve the accuracy of the newly developed μRTIS. We use MUSIC for its ability to outperform other algorithms even when using only one snapshot [15]. We also include a form of Forward-Backward Spatially Smoothed Capon (FBSS) because it has the advantage that the amplitude information of the calculated spectrum is still present while also being known for having a good angular resolution [16]. This amplitude information is lost when using any type of MUSIC because of the applied eigendecomposition of the covariance matrix. The Capon beamformer is also forward-backward spatially smoothed to ensure the localization of the present coherent sources. We also include the first real-life processed imaging results of the μRTIS while using different beamforming algorithms which were first presented in [7].

This article is structured as follows: firstly, we explain the used signal model and used algorithms in Section III. Then we clarify the methodology in Section IV. In Sections V-A and V-B we investigate respectively the influence of the Signal-to-Noise Ratio (SNR) and number of snapshots. Next, we investigate the angular separation in Section V-C followed by the investigation into the resolvability of multiple sources at a same distance for the different algorithms in Section V-D, ending with real-life results in Section V-E. Finally, in Section VI, we state the conclusion of this article.

II. μRTIS
The μRTIS is the successor of the previously developed eRTIS. The eRTIS was based on Poisson Disc-Sampling and features a 32 pseudo-random microphone array. It uses DAS

to solve the DOA problem and has a footprint of 12.2 cm by 10.8 cm. To support the use of their 3D in-air sonar sensor in mobile robotic applications in which the relatively large size could be an issue, CoSys-Lab used their experience of the eRTIS and developed a smaller version: the μRTIS. The μRTIS features 30 microphones placed in a uniform rectangular array of 5×6 and a smaller emitter, reducing its size to 5.7 cm by 4.6 cm. A side by side view of the two sensors is visible in Fig. 2. The use of a URA enables high resolution imaging techniques to be applied. Since the microphones are placed with a spacing of 3.85 mm, we can spatially localize ultrasound sources with frequencies up to 44.545 kHz. Currently we use spatially smoothed MUSIC to process the recorded data, which requires a URA and therefore was not an option with the eRTIS. Initial real-life results and a more in-depth description of the hardware and processing of the μRTIS is available in [7]. However, as previously mentioned we will in this article use a simulation of a rectangular array of 8×8 which has been spatially smoothed with subarrays of 5×5. This research environment enables us to research the influence of different parameters to the imaging results of different DOA-methods in a completely controlled environment. These simulation results can then be used to further improve the imaging of the μRTIS. We performed the simulation results on an 8×8 URA, compared to the 5×6 URA of the μRTIS, the results hereof will be better than when we apply the same techniques to real-life data. However, the same principles apply.

III. SIGNAL PROCESSING FLOW
A. SIGNAL MODEL
The following is a description of the signal model used in the simulation environment. This model will replace the received signals of step (a) in Fig. 1. Assume a 2D array of M microphones and an emitter close to the array's phase center, emitting a signal $s_e(t)$. The signal used in the simulation was a frequency sweep ranging from 20 kHz to 85 kHz. This signal gets reflected by the K point reflectors in the environment and recorded by the microphone array. We can write the m-th microphone signal as:

$$s_m(t) = \sum_{k=1}^{K} a_k \cdot s_e(t - \Delta t_{k,m}) + n(t), \qquad (1)$$

in which the time delay $\Delta t_{k,m}$ is caused by the range $r_{k,m}$ between the m-th microphone and k-th source and $n(t)$ is the noise vector. The attenuation of the signal a_k is due to a number of effects of which the most important ones are absorption by the reflector and attenuation with distance which in the free field is equal to $\frac{1}{r}$ [18]. With r being the travelled distance of the acoustic wave, which translates to $\frac{1}{2r_{k,m}}$ for our active sonar application. To focus on the performance of the beamforming algorithms we chose not to take absorption, diffraction or reverberation into account, these effects are an interesting topic for future work.
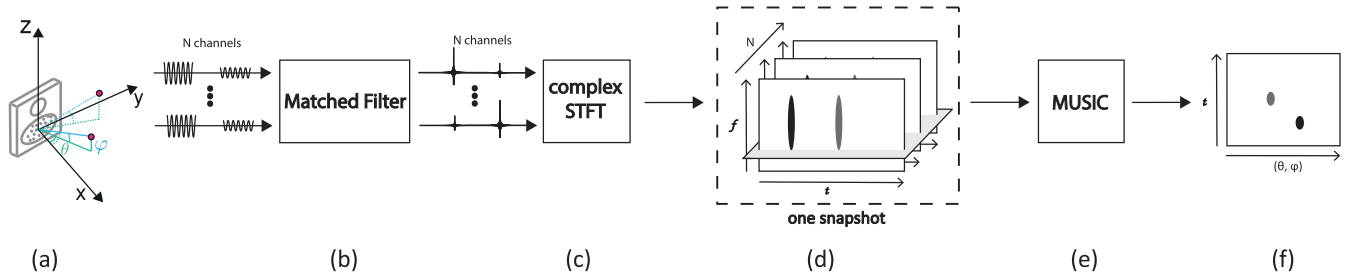
(a) (b) (c) (d) (e) (f)

FIGURE 1. Schematic overview of the steps that are executed by the signal processing flow. An adapted version of the processing flow in [17] where the coordinate system used in the simulation is defined on the eRTIS. (a) An eRTIS 3D imaging sonar performs an active measurement of the environment. (b) A matched filter is used to maximize the SNR. (c) Using complex STFT we can extract the frequency information of the recorded signal and estimate the range of the reflections (d) The one snapshot that will be used consists out of the range-(single)frequency information across multiple channels. (e) This data is used as the input of an angle estimation algorithm, in this figure MUSIC was used. (f) Finally we get a 3D image showing the location of the reflectors.
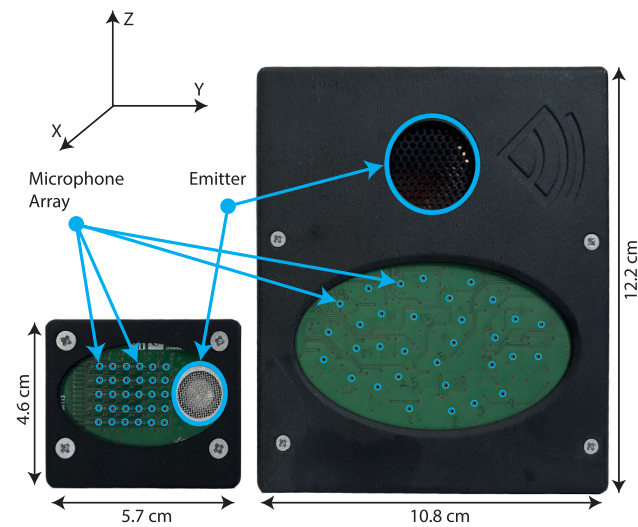

FIGURE 2. Figure from [7] showing a side by side view of the μRTIS (left) and the eRTIS (right), which are two 3D imaging sonar sensors developed by CoSys-Lab. The microphone array in the eRTIS sensor consists of a pseudo-random 32 microphone array placed using Poisson Disc-Sampling, whereas the microphone array in the μRTIS consists of a 30 microphone URA of 5 × 6.

The noise vector $n(t)$ is used to add white Gaussian noise to the model, making it possible to change the SNR of our recorded signal:

$$SNR = 10 \cdot log_{10}\left(\frac{P_{s_m(t)}}{P_{n(t)}}\right) \text{ [dB]}, \qquad (2)$$

where the $P_{s_m(t)}$ and $P_{n(t)}$ respectively are the power of the signal and the power of the background noise. Our simulated environment consists of point sources in the free field from which the reflected signals only follow one clear path to our sonar sensor. Fig. 3 illustrates the received signal of a source placed at a distance of one meter, using a spectrogram with a window size of 256 and an overlap of 240 samples.

Once the signal reaches the individual microphones, it first gets matched filtered to maximize the SNR and estimate the time of arrival (range) [19]:

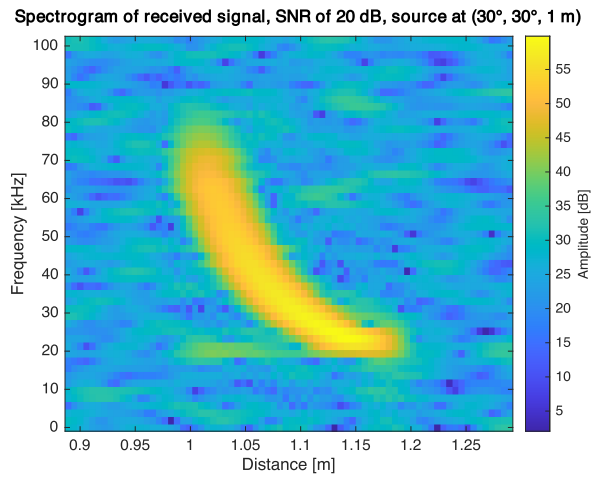$$s_m^{MF}(t) = \mathcal{F}^{-1}[\mathcal{F}(s_m(t)) \cdot \mathcal{F}^*(s_e(t))]. \qquad (3)$$

FIGURE 3. Spectrogram of the received signal with a SNR of 20 dB. The source (azimuth, elevation, range) is placed at (30°, 30°, 1 m). The received signal shows the frequency chirp starting at around 1 m and 80 kHz to 1.2 m and 20 kHz.
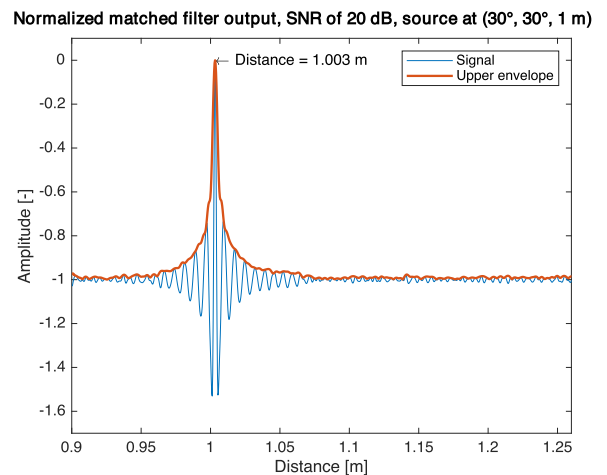

FIGURE 4. Normalized output of the matched filter for a source (azimuth, elevation, range) placed at (30°, 30°, 1 m). The maximum value is reached at a distance of 1.003 m.

Fig. 4 shows the output of the matched filter for a single microphone for the received signal from Fig. 3. In this case, where the source is placed at a distance of 1 m and not right in

front of the sensor, the distance is estimated at 1.003 m. More importantly, the Rayleigh width [20], which will be discussed further in Section IV, decreases from approximately 20 cm to around 6 mm.

This filtered signal is transformed to the time-frequency domain using a Short-Time Fourier Transform (STFT). We select the frequency bin which has the highest energy (ω_1) and use this frequency for all of our snapshots. Next, we filter out the time bins for which the total energy is less than half of the highest energy, leaving us with a limited collection of time (range) slots for a certain frequency and a certain snapshot. This results in a single snapshot $\vec{x}(\omega, t)$ for each recording, where the time t matches a distance from the sensor and ω is fixed to ω_1. Gathering multiple snapshots by measuring K times results in the matrix notation:

$$X(\omega, t) = \begin{bmatrix} \vec{x}_1(\omega_1, t) & \vec{x}_2(\omega_1, t) & \cdots & \vec{x}_K(\omega_1, t) \end{bmatrix}. \quad (4)$$

However, as previously mentioned we are limited to only one snapshot resulting in $X(\omega, t)$ being equal to $\vec{x}_1(\omega_1, t)$.

B. SIMULATION PARAMETERS

The simulated sensor consists of 64 (8 × 8) omnidirectional microphones placed in a Uniform Rectangular Array (URA). The spacing in between the microphones is 5 mm in both directions. The signal emitted by the sonar sensor is a frequency sweep, sweeping from 20 kHz to 85 kHz and a duration of 1 ms. By placing our microphones 5 mm apart.

C. DOA METHODS

Following the STFT, we have now reached step (e) in Fig. 1. The gathered snapshot serves as an input for the beamforming algorithms. We have chosen three different algorithms to calculate the DOAs of the simulated signal, i.e., DAS, MUSIC, and Capon. DAS is based on the difference of the phase of the signal as a result of the spacing between the microphones. Its spectrum is defined as

$$P_{DAS}(\theta, \phi) = X(\omega, t) \cdot A(\theta, \phi), \quad (5)$$

in which $A(\theta, \phi)$ is the steering matrix defined for azimuth θ and elevation ϕ, steering our signal into all the directions of interest which in our case is equal to the complete frontal hemisphere. We defined

$$A(\theta, \phi) = \begin{bmatrix} a(\theta_1, \phi_1) & a(\theta_2, \phi_2) & \dots & a(\theta_M, \phi_M) \end{bmatrix}, \quad (6)$$

where $a(\theta, \phi)$ is a single steering vector:

$$\psi_{Y,m} = sin(\theta) \cdot cos(\phi) \cdot p_{y,m}$$
$$\psi_{Z,m} = sin(\phi) \cdot p_{z,m}$$
$$a(\theta, \phi) = \begin{bmatrix} e^{\frac{-2\pi j}{\lambda} \cdot (\psi_{Y,1} + \psi_{Z,1})} & \cdots & e^{\frac{-2\pi j}{\lambda} \cdot (\psi_{Y,M} + \psi_{Z,M})} \end{bmatrix}. \quad (7)$$

In which λ is equal to the wavelength of the selected frequency and $[p_{y,m}, p_{z,m}]$ is equal to the location of the m-th microphone, with the azimuth and elevation defined as in Fig. 1.

Besides DAS, we also used MUSIC [8] to solve the DOA problem:

$$P_{MUSIC}(\theta, \phi) = \frac{1}{A(\theta, \phi)^H \cdot E \cdot E^H \cdot A(\theta, \phi)}, \quad (8)$$

it is based on the eigenvalue decomposition of the covariance matrix R of $X(\omega, t)$. Where $A(\theta, \phi)^H$ in (8) is equal to the hermitian transpose of $A(\theta, \phi)$. It solves the different DOAs by calculating the noise subspace E which are the eigenvectors of R belonging to the noise. The noise subspace is formed by the eigenvectors of R corresponding with eigenvalues that are smaller than the mean of all eigenvalues. The estimation results of the MUSIC algorithm are dependent by six parameters in particular: The number of array elements, spacing between the elements, number of snapshots, SNR, angle spacing, and coherency of the signals. In the course of this article we will keep the number of array elements and spacing between them constant while all of our signals are coherent. The influence of the remainder of these parameters will be the main subject of this article.

The last beamformer used is Capon, to be specific we use the version of Capon known as the Minimum Power Distortionless Response (MPDR) [21] with diagonal loading:

$$R_b = R + I_M \cdot b$$
$$w_0(\theta, \phi) = \frac{R_b^{-1} \cdot A(\theta, \phi)}{A(\theta, \phi)^H \cdot R_b^{-1} \cdot A(\theta, \phi)}$$
$$P_{MPDR}(\theta, \phi) = w_0(\theta, \phi)^H \cdot R_b \cdot w_0(\theta, \phi). \quad (9)$$

where I_M is the identity matrix of size $M \times M$, b is the value for diagonal loading and equal to -0.5 since it is best kept small and negative [22] and R^{-1} is the inverse of the estimate of the covariance matrix [23].

D. SPATIAL SMOOTHING

Since the use of a sonar sensor makes most of the signal sources coherent, spatial smoothing will be applied to overcome this limitation of the MUSIC algorithm [12], [24]. Spatial smoothing is one of the most used methods for this purpose because of its simplicity and small computational demand. Forward Spatial Smoothing (FSS) averages the signal of a microphone with that of its neighbors. This decorrelates the received signals and improves the performance of the DOA-estimators [25], [26]. Forward spatial smoothing requires a URA in the presence of K coherent sources to be at least of size $2K \times 2K$. To guarantee the localization of all present sources, forward-backward spatial smoothing will also be implemented, requiring a minimum size of $\frac{3}{2}K \times \frac{3}{2}K$ for the URA [27]. Besides MUSIC, the Capon beamformer will also be forward-backward spatially smoothed to ensure the localization of the present coherent sources. In this article, the sensor consisting of 8 × 8 microphones will be divided in subarrays consisting of 5 × 5 microphones.

IV. METHODOLOGY

We will investigate the accuracy of different DOA algorithms in function of different properties using a sonar
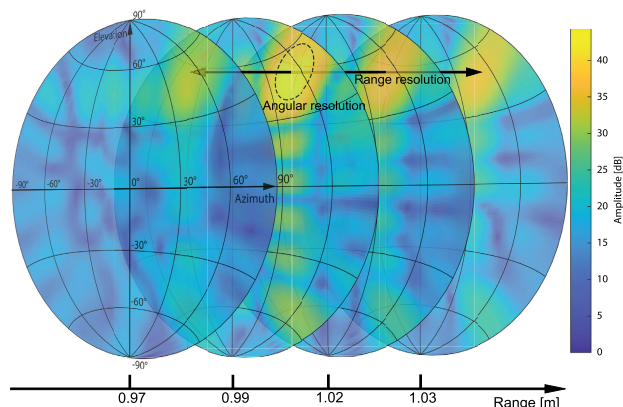
FIGURE 5. Spatial image of a 3D in-air sonar sensor using the DAS algorithm with a source (azimuth, elevation, range) placed at (45°, 45°, 1 m), a SNR of 10 dB, and one snapshot. The four slices match four sequential time slices (ranges) of the same recording. The angular resolution is defined as the surface on the sphere formed by the 3 dB area in steradians. The range resolution is defined as the distance between the 3 dB points from the different slices.

sensor in a simulation environment. The angular accuracy will be defined by the 3 dB area on the sensor's Point Spread Function (PSF) as visible in Fig. 5. Since this is an area on a sphere, it is defined in steradians (sr). The range resolution will be defined as the distance between the 3 dB points of the source. The DOA algorithms that will be compared are DAS, Capon FBSS, MUSIC, MUSIC FSS, and MUSIC FBSS. For which we selected the frequency with the highest energy to be used in (7), which was equal to 31.250 kHz. This frequency satisfies the half-wavelength criteria for our array where the distance between the microphones is equal to 5 mm.

First, we investigate the influence of the SNR of the received signal. We conduct three experiments, each experiment has one source at a distance of 1 m. Every experiment will let the SNR range from −16 dB to 20 dB and is repeated 20 times. The first experiment will have a source (azimuth, elevation, range) at (0°, 0°, 1 m), the second a source at (30°, 30°, 1 m), and the final experiment will have a source placed at (60°, 60°, 1 m). The result will be a graph displaying the average resolution and the standard deviation for every DOA algorithm. Only one snapshot will be used since we are limited to only one snapshot in real-life applications. This is a consequence of the slower speed of sound, to gather multiple measurements the signal sources should be stationary, which would limit the possibilities of the 3D sonar application.

The research into the effect of the number of snapshots to the angular and range resolution will be conducted in a similar manner as the research for the SNR. The SNR will be 3 dB to assure a clearly received signal and the number of snapshots will range from 1 to 15.

Next, we investigate how close two point sources can be placed while remaining distinguishable. Two sources will be placed at an elevation of 0° and a distance of 1 m. They will be centered around 0° azimuth and will start with a spacing of 40°, moving closer with 2° at a time. The criterion of distinguishably will be that the maximum attenuation found

in between the two sources should be higher than 3 dB. This is analogous to the Rayleigh criterion or Rayleigh width in optics [20]. This experiment will again be repeated 20 times for the different DOA algorithms with a SNR of 10 dB and only one snapshot.

Finally, we will end the simulation results with an investigation into how many sources can be placed at the same range while still being resolvable. Up to five sources will be placed at an elevation of 0° and a distance of 1 m. They will be centered around 0° azimuth with a spacing of 40° in between, again with a SNR of 10 dB and only one snapshot. This experiment will be repeated 20 times for the different DOA algorithms. The output will be the sum of the spatial images for every range averaged over the 20 repetitions.

We end our results with a real-life imaging run of the μRTIS. We attached the μRTIS to a Pioneer 3-DX, a small radio guided robot, that was driven through a narrow corridor as is visible in Fig. 8. We captured data at 10 Hz and calculated acoustic images using DAS, Capon, and MUSIC FBSS.

V. RESULTS AND DISCUSSION

A. SNR

Fig. 6 shows the angular resolution in function of the SNR for a source placed at (0°, 0°, 1 m), (30°, 30°, 1 m), and (60°, 60°, 1 m). The SNR ranges from −16 dB to 20 dB. It can be seen that the angular resolution of the MUSIC algorithm applied to a sonar sensor is very dependent on the SNR, this is the most clear for the case where the source is placed (60°, 60°, 1 m). However, when the SNR is higher than 0 dB in the latter case, the resolution appears to be stable. The same trend can be seen for Capon FBSS, MUSIC FSS, and MUSIC FBSS, all be it less severe. The angular resolution for these algorithms is in the worst case around 0.025 sr, compared to 0.3 sr for the MUSIC algorithm. These algorithms also become stable around −10 dB. DAS is the worst performer, the angular resolution for every case is quite stable but the highest of all algorithms when the SNR increases.

Fig. 6 further shows a deterioration of the angular resolution when the source is placed at higher angles. This is the most visible for the DAS algorithm, for which the angular resolution rises from 0.04 sr when the source is placed at (0°, 0°, 1 m), to 0.12 sr for a source placed at (60°, 60°, 1 m). This deterioration is an immediate effect of the decreased aperture of the sonar sensor. The aperture is proportionate to the cosine, meaning it is halved when the signal originates from −60° or 60°. The angular resolution of all beamformers in this situation worsens.

Fig. 7 shows the range resolution in function of the SNR for a source placed at (0°, 0°, 1 m), (30°, 30°, 1 m), and (60°, 60°, 1 m). The range resolutions for all algorithms are quite stable for all cases with an exception for MUSIC and MUSIC FSS when the source is placed at (0°, 0°, 1 m). For these cases the range resolutions worsens for higher SNRs. Overall the differences in range resolutions between the tested algorithms is negligible. When the source is not placed right in front of
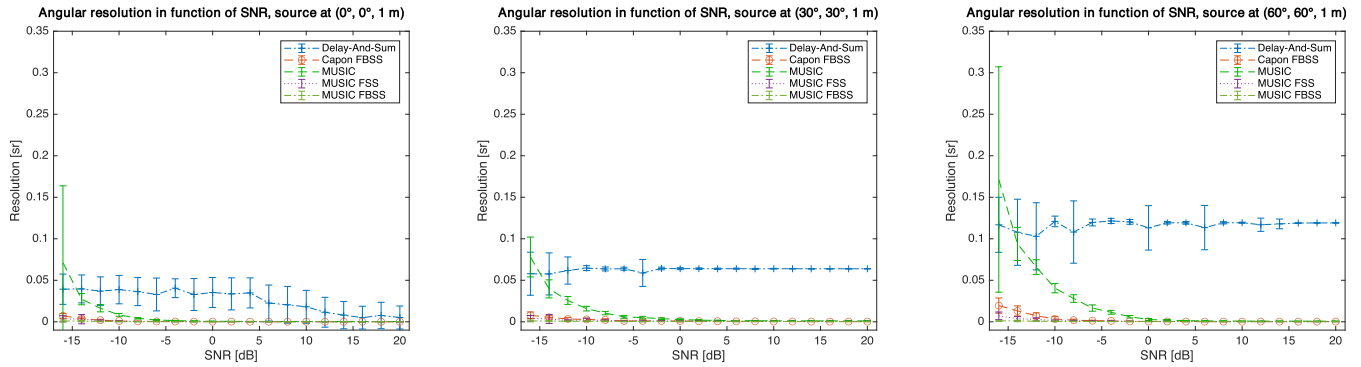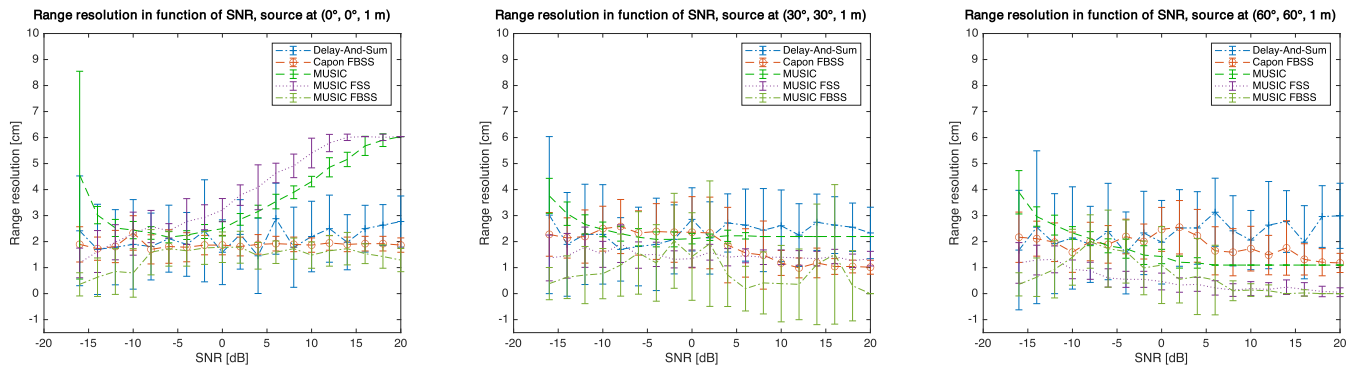
FIGURE 6. Angular resolution in function of SNR with five different DOA-algorithms and one snapshot, a source (azimuth, elevation, range) is placed at (0°, 0°, 1 m), (30°, 30°, 1 m), and (60°, 60°, 1 m).

FIGURE 7. Spatial resolution in function of SNR with five different DOA-algorithms and one snapshot, a source (azimuth, elevation, range) is placed at (0°, 0°, 1 m), (30°, 30°, 1 m), and (60°, 60°, 1 m).

FIGURE 8. Setup of the real-life imaging run of the μRTIS. The μRTIS is attached to a Pioneer 3-DX and was driven through a narrow corridor of approximately two meters wide at our university.

the sensor all sensors perform quite alike with overlapping standard deviations that do not exceed 6 cm. Which in most use cases is sufficient.

B. NUMBER OF SNAPSHOTS

Fig. 9 shows the angular resolution is not dependent of the number of snapshots in a low snapshot scenario except for

the original MUSIC algorithm. This effect is most visible in the detailed view of the Capon FBSS and MUSIC algorithms, in particular when the source is placed at (60°, 60°, 1 m). MUSIC's mean angular resolution evolves from 0.003 sr for one snapshot to 0.001 sr for five snapshots whereas the spatially smoothed algorithms remain stable. This is a very positive effect since the biggest threshold of using MUSIC in a sonar application is the difficulty of recording multiple snapshots. Fig. 9 further shows the immense gain in resolution compared to DAS. The resolutions for the smaller angles are all stable, but whereas the spatially smoothed MUSIC algorithms and Capon FBSS remain under 0.0016 sr, DAS has a mean resolution of nearly 0.025 sr when the source is placed at (0°, 0°, 1 m) and 0.125 sr for a source placed at (60°, 60°, 1 m). The bottom row of Fig. 9 further reveals that MUSIC FBSS and MUSIC FSS perform slightly better than Capon FBSS, their angular resolutions are the most stable for all angles. This is important for 3D sonar applications that scan the entire frontal hemisphere.

The range resolution results visible in Fig. 10 also show no connection between the number of snapshots and the range resolution. The maximum resolution of about 7 cm is the consequence of the preprocessing. To avoid calculating every possible partial spatial spectrum, the time slots get filtered according to the maximum amplitude detected in the entire spectrum. When the total amplitude of a time slot is lower than half the maximum value, its amplitude is neglected.
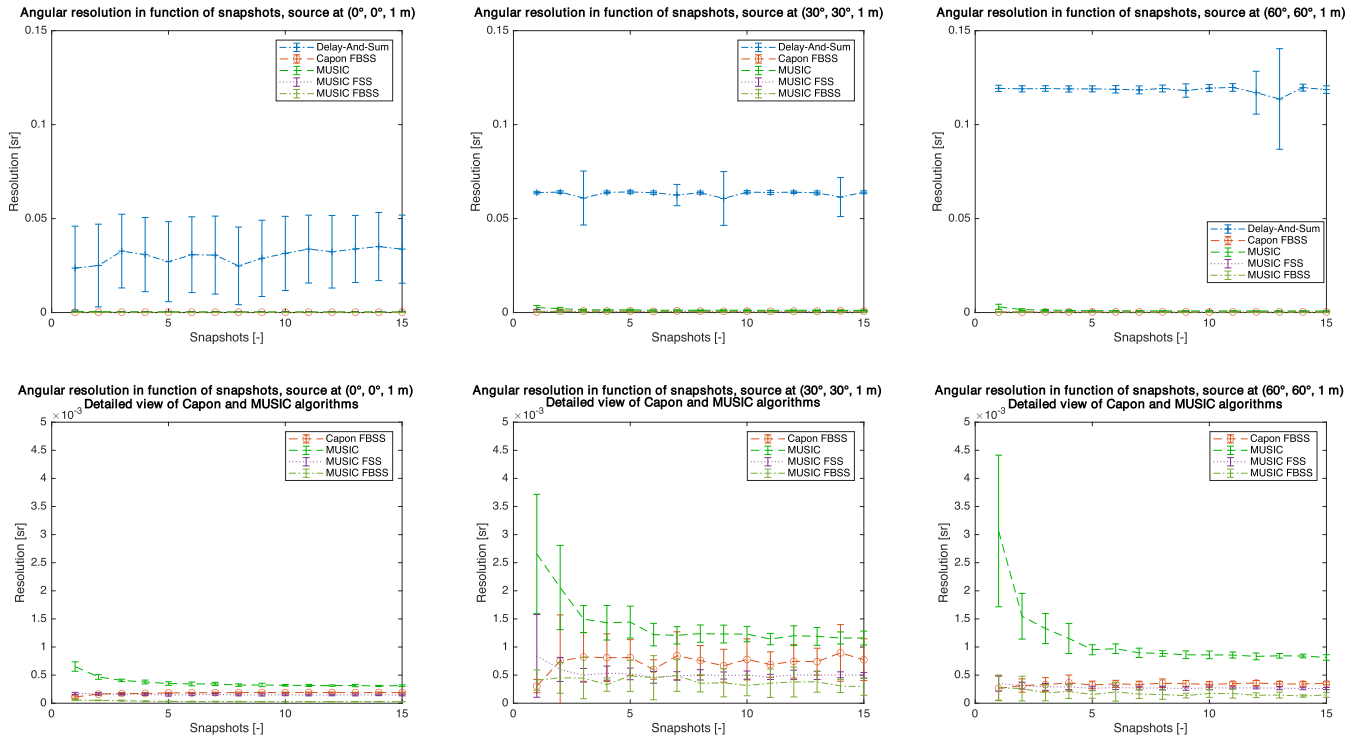
FIGURE 9. Angular resolution in function of snapshots with five different DOA-algorithms and a SNR of 3 dB, a source (azimuth, elevation, range) is placed at (0°, 0°, 1 m), (30°, 30°, 1 m), and (60°, 60°, 1 m).

FIGURE 10. Spatial resolution in function of snapshots with five different DOA-algorithms and a SNR of 3 dB, a source (azimuth, elevation, range) is placed at (0°, 0°, 1 m), (30°, 30°, 1 m), and (60°, 60°, 1 m).
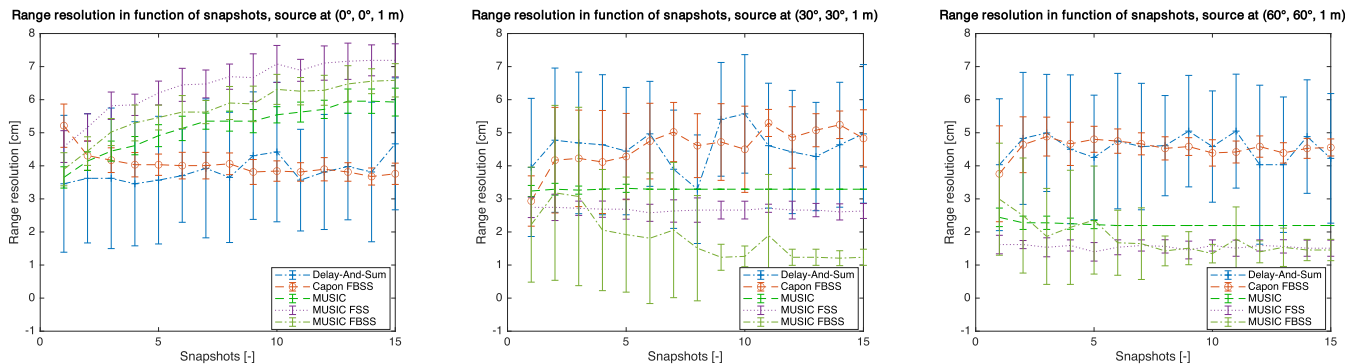
Overall, DAS and Capon FBSS perform the most stable for all the cases, whereas it can be seen that the MUSIC algorithms are very dependent on the location of the source, improving slightly as the source moves away from the center of the sensor.

C. SPACING

Fig. 11 shows the maximum attenuation of the acquired spectrum between two point sources. This provides a good indication of how close two sources can be placed while still remaining distinguishable. The 3 dB line will be used as a limit, when the maximum attenuation is lower than the 3 dB line it is fair to say it is no longer possible to distinguish

the two sources from each other. MUSIC performs the worst out of the different algorithms, its maximum attenuation does not pass 8 dB and at 25° it crosses the 3 dB line. It is even outperformed by DAS which is able to distinguish the two sources with a spacing of 24° with attenuations well over 10 dB. However, it is important to mention that this figure only displays the maximum attenuation between two sources. In Fig. 12 it can be seen that DAS shows extra peaks in between the real sources which can wrongly be identified as incoming signals. This effect can also be seen with the MUSIC algorithm, all be it less severe.

Fig. 11 further shows the effect of the spatial smoothing, the spatial smoothed algorithms (Capon FBSS, MUSIC FSS, and MUSIC FBSS) all perform quite alike. They are able to
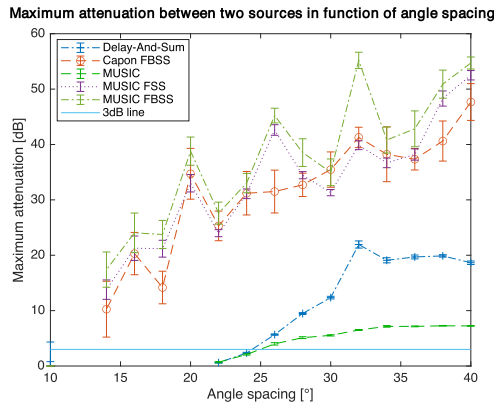
FIGURE 11. Maximum attenuation between two sources in function of angle spacing with five different DOA-algorithms. The elevation and range of the sources is kept constant at 0° and 1 m, while the azimuth spacing varies from 40° to 10°, centered around 0°. A SNR of 10 dB and one snapshot is used.

clearly distinguish the two sources up until a separation of 14° with a mean attenuation that stays above 10 dB.

D. RESOLVABILITY

Fig. 12 depicts the different spectra for the five used DOA-algorithms when placing multiple reflectors at the same distance. The spectrum of DAS resolves all the sources clearly but in between the real sources, the limited dynamic range could lead to a wrongly identified source. These artefacts are not present with the remaining algorithms. For less than five sources Capon FBSS resolves the sources correctly.

The dynamic range of the spectrum is around 60 dB, making Capon FBSS have the highest dynamic range along with MUSIC FBSS. However, when five sources are placed at 1 m all at the same elevation, Capon and the MUSIC algorithms fail to correctly resolve the sources. This is a consequence of the coherency of the sources. Due to the fact that all of the sources are placed directly in front of the sonar sensor at an elevation of 0° our rectangular array will act as a linear one. In that case we only have four subarrays in the plane with an elevation of 0° and are capable of resolving up to four sources [12]. This results in the sources not being resolved correctly. Both Capon and the MUSIC algorithms seem to locate only four or three sources respectively, most of them located at the wrong locations. The same effect can be seen with DAS, where the Rayleigh criterion is no longer satisfied merging the responses of the four outermost sources into two erroneous responses. This is not a big issue in real-life performance since it is very rare to have five or more coherent sources falling within the Rayleigh width (approximately 6 mm in Fig. 4) of the cross correlation function of the emitted chirp.

MUSIC performs well except for a very limited dynamic range when there are more than one source present. MUSIC FSS and MUSIC FBSS perform alike. However, MUSIC FBSS performs better when less than five sources are present, for these situations the resolution and dynamic range of MUSIC FBSS are better.

As previously mentioned in Section V-A, the aperture of the sensor dramatically worsens for angles higher than 60°. This is visible when four sources are placed. In this case,
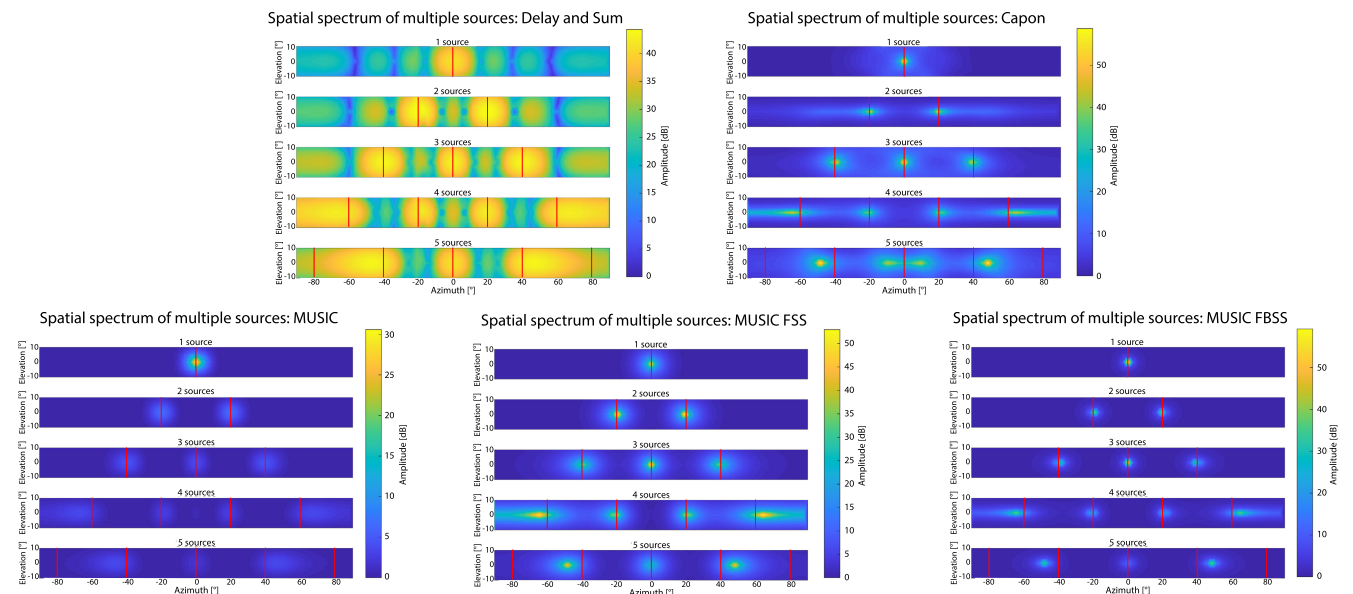

FIGURE 12. Spectra of five different DOA-algorithms with multiple sources centered around 0° azimuth with a spacing of 40°. The sources are placed at a distance of 1 m and a SNR of 10 dB. In the case of one source, the source is placed at 0° azimuth. When two sources are present, the sources are placed at −20° and 20° azimuth. In the case of three sources, the sources are placed at −40°, 0°, and 40° azimuth. When four sources are present, the sources are placed at −60°, −20°, 20°, and 60° azimuth. In the final case of five sources, the sources are placed at −80°, −40°, 0°, 40°, and 80° azimuth. The real locations of the sources are indicated by vertical red lines.
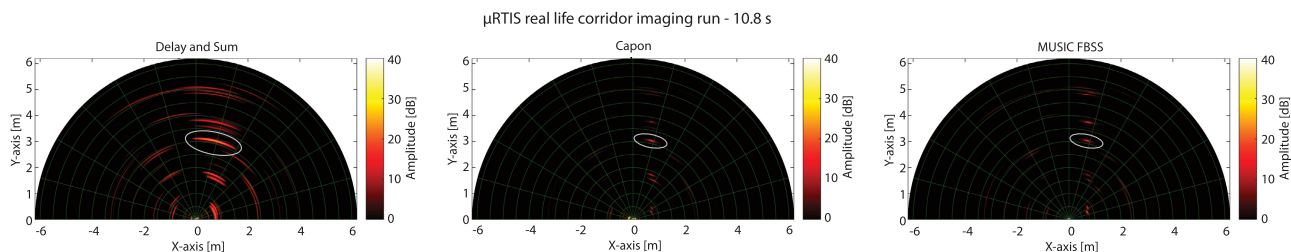
FIGURE 13. Real-life imaging results of the μRTIS. The recorded data was processed using Delay-And-Sum, Capon, and MUSIC FBSS. The μRTIS was attached to a small radio guided robot and driven through a narrow corridor. This figure is a snapshot at 10.8 s, the full run is available at [28].

FIGURE 14. Real-life imaging results of the μRTIS. The recorded data was processed using Delay-And-Sum, Capon, and MUSIC FBSS. The μRTIS was attached to a small radio guided robot and driven through a narrow corridor. This figure is a snapshot at 38.3 s, the full run is available at [28].
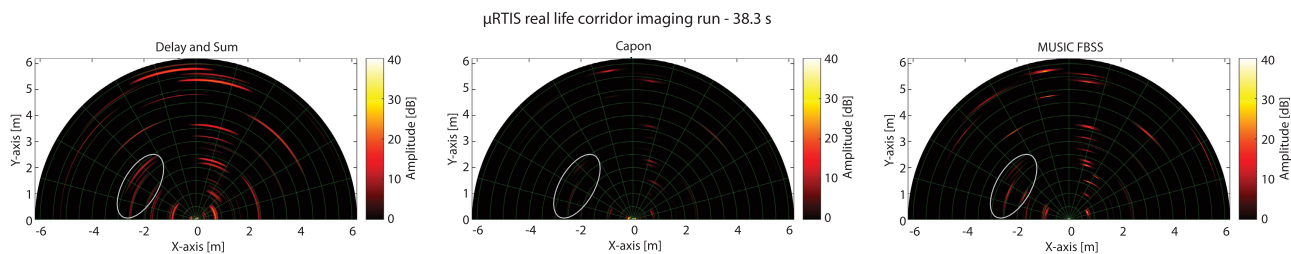
the PSFs of the sources at $-60°$ and $60°$ are wider than those of the other sources. The resolved sources in this situation also appear to have a slight offset that worsens for sources placed at higher angles as an effect of the decreased aperture of the sensor relative to these sources, which was also discussed in Section V-A.

In Section V-B it was seen that Capon FBSS had one of the lowest angular resolutions when there is one source present, whereas MUSIC FBSS performed the worst. However, Fig. 12 shows that the MUSIC algorithms, especially MUSIC FBSS, have lower angular resolutions when more than one source is present.

E. REAL-LIFE IMAGING RESULTS

The μRTIS was attached to a small radio guided robot and driven through a narrow corridor at the university of about 2 m wide. The data was captured at 10 Hz and the acoustic images were calculated using DAS, Capon, and MUSIC FBSS, using the processing scheme described in 1. A video of the entire dataset is available at [28]. Two snapshots from this video were taken to discuss the differences between the acquired results. Fig. 13 and Fig. 14 are snapshots taken at 10.8 s and 38.3 s respectively. In each figure a gray ellipse was drawn around a reflection for reference.

Fig. 13 shows a clear improvement in the imaging results of Capon and MUSIC FBSS compared to DAS. The point spread functions of the solved reflections are much narrower. The green lines are drawn every $15°$, meaning the PSFs of the imaging results of DAS span over an area of about $30°$ compared to about $7.5°$ when Capon or MUSIC FBSS is used. All though the PSFs are narrower, the dynamic range slightly

worsens, it can be seen that Capon and MUSIC FBSS solve the sources with a maximum peak that is lower than 40 dB, which is the maximum reached with DAS.

Fig. 14 on the other hand shows a clear difference between Capon and MUSIC FBSS. In this figure four reflections are indicated which are clearly resolved using DAS and MUSIC FBSS. However, Capon only resolves two of them and with a magnitude which is very low compared to the other two algorithms. Since these reflections are at the outer rim of the sensor, centered at around $-60°$, it can also be seen the PSFs are wider compared to the results in Fig. 13. The same effect was also previously discussed during the simulations.

VI. CONCLUSION

The simulation results revealed one of the biggest bottlenecks of sonar sensors: the need for a high dynamic range. Reflectors that are placed at higher angles suffer from harsh attenuation of the signal strength due to decrease in aperture of the sensor. Sources at a distance of 1 m are easily solved with SNR higher than 0 dB.

Unexpectedly, the SNR and the number of snapshots no longer have a high influence to the angular and spatial resolution of the different MUSIC algorithms. This effect is especially true for the number of snapshots when spatial smoothing is applied, which is a very positive observation since we are limited in the number of snapshots to one due to time constraints in our measurement process.

From the different algorithms that were compared, MUSIC FSS and MUSIC FBSS have the best angular and spatial resolution when the sources are scattered (not placed at $(0°, 0°, 1$ m$)$). Their angular resolution also stays stable

when the location of the incoming signal changes between 0° and 60°. In contrast to DAS which always has an angular resolution that is up to almost 100 times greater than the other algorithms. The angular resolution of MUSIC and Capon FBSS in function of the SNR behave quite alike, with Capon FBSS behaving the best of both.

The range resolutions of DAS, Capon FBSS, and MUSIC FBSS all behave similarly for the different sources, in most cases they do not pass 6 cm which is fine for most 3D sonar applications.

Spatial smoothing should be used when it is necessary to solve multiple reflectors that are placed closely together, the spatially smoothed algorithms outperform DAS and MUSIC with more than 10°. However, it should be noticed that Capon FBSS does show some artefacts that could be wrongly identified as sources.

Overall, the spatially smoothed algorithms performed the most stable in the presence of one source. The angular resolutions are the lowest and most stable of the compared algorithms. The great advantage of Capon is that the magnitude information of the incoming sources is preserved. If this information is important for the application, a form of Capon should be used. However, it is important to use a sufficient spatial smoothing scheme to suppress the artefacts that are visible when multiple sources are present. In these cases the spatially smoothed MUSIC algorithms perform better.

Another factor of the MUSIC algorithm that not appears to be an issue in our sonar application, is the coherency of the incoming signals when spatial smoothing is not used. When less than five sources are placed at the same distance and therefore all the signals coherent, the original MUSIC algorithm is still able to resolve the different DOAs. It is of importance to notice that MUSIC FSS and MUSIC FBSS do have a higher dynamic range and better resolution in this situation.

These results prove the feasibility of the μRTIS in a simulated environment and this was further illustrated with a real-world system. The real-life imaging results confirmed some of the simulation results such as the ability of MUSIC and Capon to greatly improve the imaging results. It further showed that DAS and MUSIC were able to correctly identify all the sources that were present whereas Capon only identified half of them in a certain situation. These sources that were resolved using Capon also had a lower amplitude and were barely visible in the imaging results. This effect was also visible in the simulation results when multiple sources were placed at the outer rim of the sensor.

REFERENCES

[1] A. Filgueira, H. González-Jorge, S. Lagüela, L. Díaz-Vilariño, and P. Arias, "Quantifying the influence of rain in LiDAR performance," *Measurement*, vol. 95, pp. 143–148, Jan. 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0263224116305577

[2] T. G. Phillips, N. Guenther, and P. R. McAree, "When the dust settles: The four behaviors of LiDAR in the presence of fine airborne particulates," *J. Field Robot.*, vol. 34, no. 5, pp. 985–1009, Aug. 2017, doi: 10.1002/rob.21701.

[3] B. Siciliano and O. Khatib, *Springer Handbook of Robotics*. Berlin, Germany: Springer, 2008. [Online]. Available: http://link.springer.com/10.1007/978-3-540-30301-5

[4] R. Kerstens, D. Laurijssen, and J. Steckel, "Low-cost one-bit MEMS microphone arrays for in-air acoustic imaging using FPGA's," in *Proc. IEEE Sensors*, Oct. 2017, pp. 1–3. [Online]. Available: http://ieeexplore.ieee.org/document/8234087/

[5] R. Kerstens, D. Laurijssen, and J. Steckel, "ERTIS: A fully embedded real time 3D imaging sonar sensor for robotic applications," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May 2019, pp. 1438–1443. [Online]. Available: https://ieeexplore.ieee.org/document/8794419/

[6] J. Steckel, A. Boen, and H. Peremans, "Broadband 3-D sonar system using a sparse array for indoor navigation," *IEEE Trans. Robot.*, vol. 29, no. 1, pp. 161–171, Feb. 2013. [Online]. Available: http://ieeexplore.ieee.org/document/6331017/

[7] T. Verellen, R. Kerstens, D. Laurijssen, and J. Steckel, "Urtis: A small 3D imaging sonar sensor for robotic applications," in *Proc. ICASSP-IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 4801–4805. [Online]. Available: https://ieeexplore.ieee.org/document/9053536/

[8] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986. [Online]. Available: http://ieeexplore.ieee.org/document/1143830/

[9] J. H. Cozzens and M. J. Sousa, "Source enumeration in a correlated signal environment," *IEEE Trans. Signal Process.*, vol. 42, no. 2, pp. 304–317, Feb. 1994. [Online]. Available: http://ieeexplore.ieee.org/document/275604/

[10] H. Krim and J. G. Proakis, "Smoothed eigenspace-based parameter estimation," *Automatica*, vol. 30, no. 1, pp. 27–38, Jan. 1994. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/0005109894902267

[11] D. B. Williams, "Detection: Determining the number of sources," in *Wireless, Networking, Radar, Sensor Array Processing and Nonlinear Signal Processing* (The Digital Signal Processing Handbook, Second Edition). Boca Raton, FL, USA: CRC Press, 2009, pp. 8-1–8-10.

[12] T.-J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-33, no. 4, pp. 806–811, Aug. 1985. [Online]. Available: http://ieeexplore.ieee.org/document/1164649/

[13] M. Wang and B. Xiao, "Separation of cogerent multi-path signals with improved MUSIC algorithm," in *Proc. Int. Conf. Comput. Inf. Sci. (ICCIS)*, Oct. 2011, pp. 992–995. [Online]. Available: http://ieeexplore.ieee.org/document/6086369/

[14] Q. Chen and R. Liu, "On the explanation of spatial smoothing in MUSIC algorithm for coherent sources," in *Proc. Int. Conf. Inf. Sci. Technol.*, Mar. 2011, pp. 699–702. [Online]. Available: http://ieeexplore.ieee.org/document/5765342/

[15] C. Degen, "On single snapshot direction-of-arrival estimation," in *Proc. IEEE Int. Conf. Wireless Space Extreme Environ. (WiSEE)*, Oct. 2017, pp. 92–97. [Online]. Available: http://ieeexplore.ieee.org/document/8124899/

[16] L. C. Godara, "Application of antenna arrays to mobile communications. II. Beam-forming and direction-of-arrival considerations," *Proc. IEEE*, vol. 85, no. 8, pp. 1195–1245, Aug. 1997. [Online]. Available: http://ieeexplore.ieee.org/document/622504/

[17] R. Kerstens, T. Verellen, and J. Steckel, "Comparing adaptations of MUSIC beamforming for 3D in-air sonar applications with coherent source-spaces," in *Proc. CD 8th Berlin Beamforming Conf.*, Mar. 2020, pp. 1–12.

[18] F. A. Everest and K. C. Pohlmann, *Master Handbook of Acoustics*, 5th ed. New York, NY, USA: McGraw-Hill, 2009.

[19] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-24, no. 4, pp. 320–327, Aug. 1976. [Online]. Available: http://ieeexplore.ieee.org/document/1162830/

[20] W. M. Saslow, *Electricity, Magnetism, and Light*. Amsterdam, The Netherlands: Elsevier, 2002. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/B9780126194555X50001

[21] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. New York, NY, USA: Wiley, Mar. 2002. [Online]. Available: http://doi.wiley.com/10.1002/0471221104

[22] F. Vincent and O. Besson, "Steering vector uncertainties and diagonal loading," in *Proc. Process. Workshop, Sensor Array Multichannel Signal*, Jul. 2004, pp. 327–331. [Online]. Available: http://ieeexplore.ieee.org/document/1502963/

[23] L. B. Fertig, "Statistical performance of the MVDR beamformer in the presence of diagonal loading," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop. (SAM)*, Jan. 2000, pp. 77–81. [Online]. Available: http://ieeexplore.ieee.org/document/877972/

[24] J. Wen, B. Liao, and C. Guo, "Spatial smoothing based methods for direction-of-arrival estimation of coherent signals in nonuniform noise," *Digit. Signal Process., Rev. J.*, vol. 67, pp. 116–122, Aug. 2017.

[25] Y. Yang, C. Wan, C. Sun, and Q. Wang, "DOA estimation for coherent sources in beamspace using spatial smoothing," in *Proc. 4th Int. Conf. Inf., Commun. Signal Process., 4th Pacific Rim Conf. Multimedia. Proc. Joint*, vol. 2, Dec. 2003, pp. 1028–1032. [Online]. Available: http://ieeexplore.ieee.org/document/1292615/

[26] B. Widrow, K. Duvall, R. Gooch, and W. Newman, "Signal cancellation phenomena in adaptive antennas: Causes and cures," *IEEE Trans. Antennas Propag.*, vol. AP-30, no. 3, pp. 469–478, May 1982. [Online]. Available: http://ieeexplore.ieee.org/document/1142804/

[27] H. Yi and X. Zhou, "On 2D forward-backward spatial smoothing for azimuth and elevation estimation of coherent signals," in *Proc. IEEE Antennas Propag. Soc. Int. Symp.*, vol. 2, Aug. 2005, pp. 80–83. [Online]. Available: https://ieeexplore.ieee.org/document/1551940/

[28] J. Steckel. (2019). *μRTIS: Comparison of DOA Algorithms*. [Online]. Available: https://www.youtube.com/watch?v=FkSgGHNXEVQ&ab_channel=CoSys-Lab

THOMAS VERELLEN received the M.Sc. degree in electronics and ICT engineering technology from the University of Antwerp, Belgium, in 2019, where he is currently pursuing the Ph.D. degree. Following his internship, he did his master's program at the Constrained Systems Laboratory, University of Antwerp. As a member of the Flanders Make Strategic Research Centre, he researches on the adaptation of ultrasound in the field of condition monitoring and predictive maintenance.

ROBIN KERSTENS received the B.Sc. degree in applied engineering electronics and ICT and the M.Sc.Eng. degree in electronics and ICT engineering technology from the University of Antwerp, Antwerp, Belgium, in 2014. He is currently pursuing the Ph.D. degree in advanced array signal processing for in-air sonar, after a short spell in the automotive industry designing pressure sensors. His bachelor's thesis was focused on RADAR signal processing and the master's thesis was focused on sonar beamforming. His research interests include signal processing, wave propagation, acoustics, machine learning, and algorithm development.

JAN STECKEL (Member, IEEE) received the degree in electronic engineering from the Karel de Grote University College, Hoboken, in 2007, and the Ph.D. degree from the Active Perception Laboratory, University of Antwerp, in 2012. His Ph.D. dissertation titled Array Processing for In-Air Sonar Systems-Drawing Inspirations From Biology. During this period, he has developed state-of-the art sonar sensors, both biomimetic and sensor-array based. During his postdoctoral period, he was an Active Member of the Centre for Care Technology, University of Antwerp, where he was in charge of various healthcare-related projects concerning novel sensor technologies. Furthermore, he pursued industrial exploitation of the patented 3-D array sonar sensor which was developed in collaboration during his Ph.D. In 2015, he became a tenure track Professor at the Constrained Systems Laboratory, University of Antwerp, where he researches on sensors, sensor arrays, and signal processing algorithms using an embedded, constrained systems approach.

• • •