Sound-source Localisation using a Microphone-array for NUbots Interim Report

Clayton Carlon, C3327986

June 13, 2023

Problem and Background

To locate a sound-source akin to that of humans is a long sought after ability in technology.
The problem is: *to develop a sound-source-localisation technique on an array of microphones using a combination of electronics, embedded computing, and signal-processing.*

Ideally, the system will be used on the robots in the NUbots team to locate acoustic events on the field and possibly interact with humans.

Scope

Essentially, the system should:

- estimate the three-dimensional direction,
- locate a reasonably distinct sound in a moderate environment, e.g. a whistle, a lone voice, a loud thud,
- and handle the noise from the motors.

Ideally, it can:

- estimate the distance along with direction,
- track the location over time using e.g. a Kalman filter,
- locate multiple simultaneous sources,
- spatially filter the localised sound,
- and work well enough in noisy and reverberant environments, e.g. in a hall full of people.

Main Methods

Only classical signal-processing methods are reviewed.
Three main methods in the literature are:

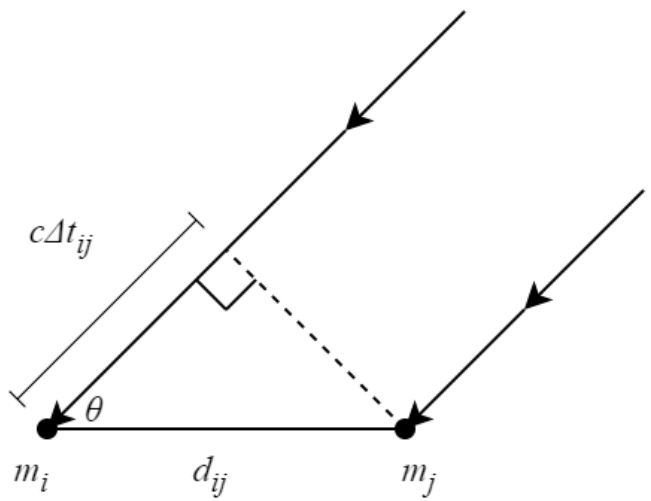
- time-difference of arrival (TDOA),
- beamforming,
- and multiple-signal-classification (MUSIC).

Time-Difference of Arrival

A simple yet proven way to estimate the direction or the location is to measure the time-delay or TDOA between microphones. For example, a basic formula is:

$$\theta = \sin\left(\frac{c\Delta t_{ij}}{d_{ij}}\right) \tag{1}$$

where c is the speed of sound, and d_{ij} is the displacement between the microphones, and assuming that the source is in the far field.



Cross-Correlation

The most common way to compute the TDOA is to find the peak of the cross-correlation of two signals x_i and x_j:

$$R_{ij}(\tau) = \sum_{n=0}^{N-1} x_i[n]x_j[n-\tau] = \mathfrak{F}^{-1}\left(X_i[k]X_j[k]^*\right) \tag{2}$$

where X_i and X_j are the Fourier transforms of two signals.

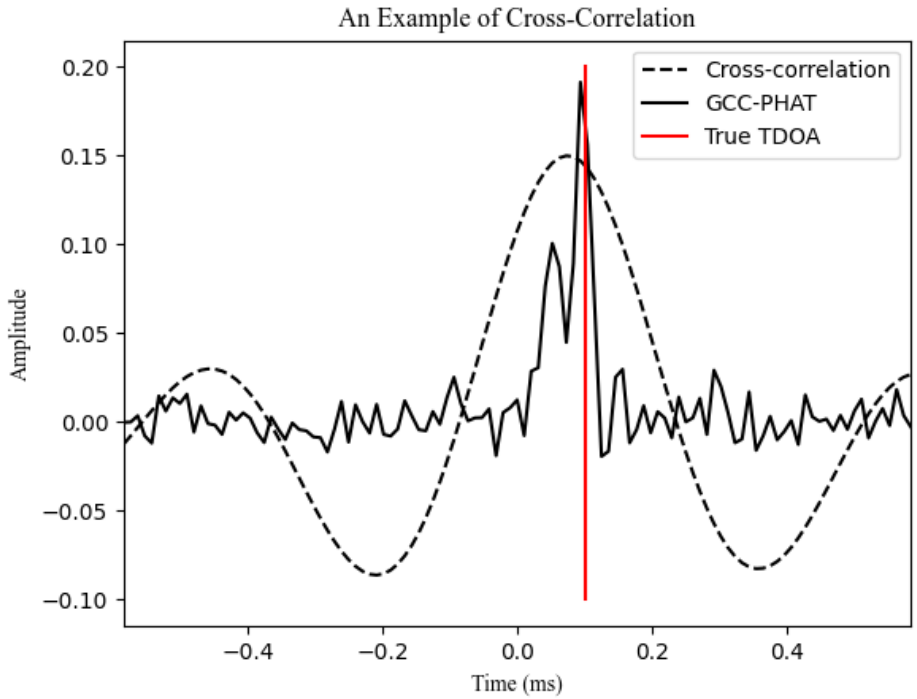
The generalised cross-correlation (GCC) is:

$$R_{ij}^{(w)}(\tau) = \mathfrak{F}^{-1}\left(\psi[k]X_i[k]X_j[k]^*\right) \tag{3}$$

where $\psi[k]$ is a spectral weighting.

A common weighting is to whiten the whole spectrum which is known as the phase-transform (PHAT) [1], [4]. This is known to improve accuracy especially against reverberation.

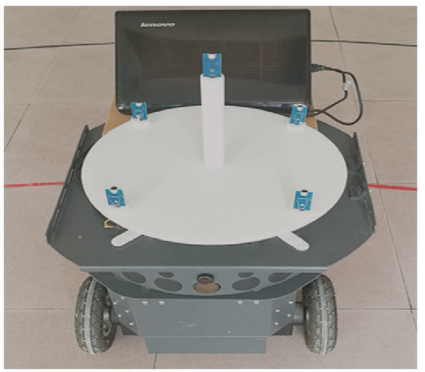
Generalised Cross-Correlation and Phase Transform



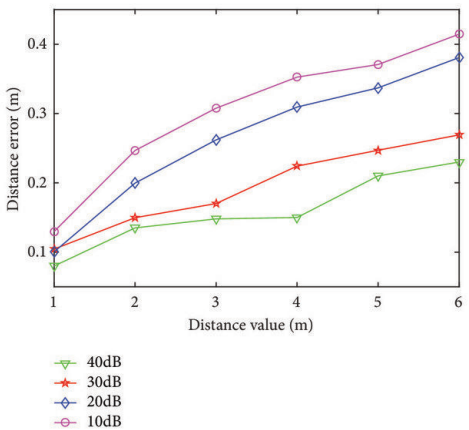
- The resolution can be improved with up-sampling.
- There is a peak at zero for narrowband signals and high noise.

Literature-example

Chen & Xu (2019), *A Sound Source Localization Device Based on Rectangular Pyramid Structure for Mobile Robot* [2]



- Estimated distances up to 6 m as well as direction.
- Used a slightly modified PHAT to deal with reverberation.
- At worst, the error in distance was 0.4 m for an SNR of 10 dB.
- The error in azimuth was within 1.5 deg.



[2]

Beamforming

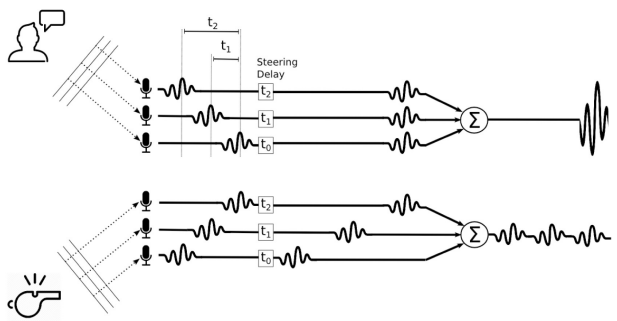
Is a very common signal-processing technique. Signals from many microphones are delayed such that they constructively interfere given a direction or location and form a beam.
The basic beamformer is the delay-and-sum kind [1].

$$y_{\vec{u}_0}[t] = \sum_{m=1}^{M} x_m[t - \tau_m(\vec{u}_0)] \tag{4}$$

where $\tau(\vec{u}_0)$ is the time-delay at the m-th microphone corresponding at the direction $\vec{u}_0$.

Given a grid of directions, this output can yield an energy-map as:

$$E_{\vec{u}_0} = \sum_{t=1}^{T} y_{\vec{u}_0}[t]^2 \tag{5}$$

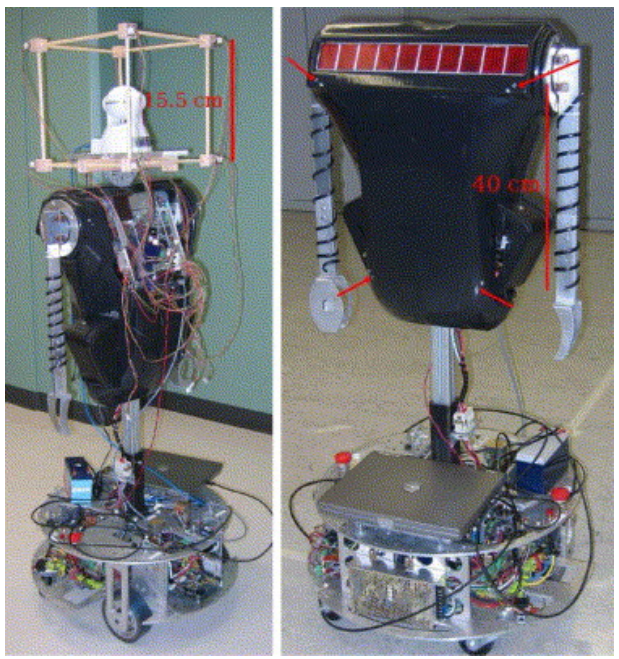


[4]

Literature-example

Valin et al. (2007), *Robust Localization and Tracking of Simultaneous Moving Sound Sources Using Beamforming and Particle Filtering* [5]

- Whitened the spectrum.
- Used a spherical search-grid of 2562 points which gave a resolution of 2.5 deg.
- Applied a particle filter.
- Switched between a coarse and a fine search grid to improve the resolution.
- At worst, the errors in azimuth and elevation were 1.10 and 1.44 deg.



[5]

Multiple-Signal Classification

Separates the space of signals into two subspaces, namely one for signals and another for noise. This is done through the eigen-decomposition of the sampled covariance matrix $R[f] \in \mathbb{C}^{N \times N}$. There are broadly three kinds, namely:

- standard eigen-value decomposition (SEVD),
- generalised eigen-value decomposition (GEVD),
- and generalised singular-value decomposition (GSVD).

Tends to have very high-resolution and robustness to noise and interference but is very computationally heavy.

Generalised Eigen-Value Decomposition

In GEVD, the eigen-decomposition is done on:

$$K^{-1}[f]R[f] = Q[f]\Lambda[f]Q^{-1}[f] \tag{6}$$

where $\Lambda[f]$ is a diagonal matrix of the M eigenvalues $\lambda_m[f]$, and each column of $Q[f]$ is a corresponding eigenvector $q_m[f]$ [3].

This matrix of eigenvectors is often split into two subspaces, $Q[f] = [Q_s[f]|Q_n[f]]$. Also, $K[f]$ is a freely chosen matrix but is often computed as $N[f]N^*[f]$ where $N[f]$ is the frequency-domain noise recorded when there are no signals.

Like beamforming, an energy-map is drawn:

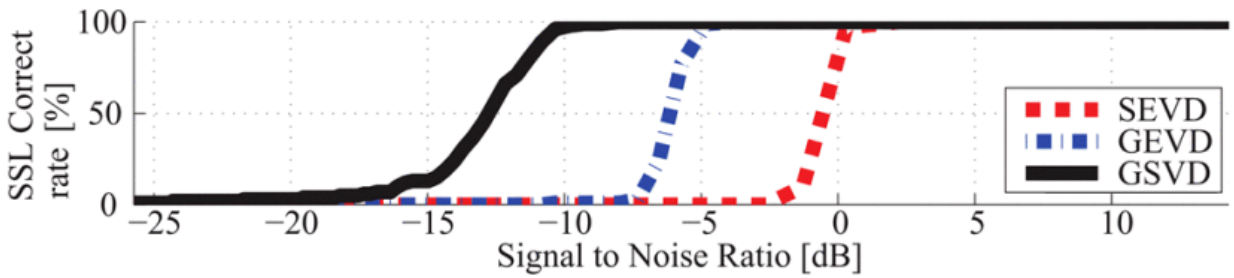
$$P(\theta_0, \phi_0)[f] = \frac{|A^*(\theta_0, \phi_0)A(\theta_0, \phi_0)|}{\sum_{m=\tilde{N}+1}^{M} |A^*(\theta_0, \phi_0)q_m[f]|} \tag{7}$$

where $\tilde{N}$ is the number of sources considered, and $A(\theta_0, \phi_0) \in \mathbb{C}^{M\times 1}$ is the steering vector.

Literature-example

Nakamura et al. (2007), *Real-Time Super-Resolution Sound Source Localization for Robots* [3]

- Proposes GSVD as computationally lighter.
- The average error in the azimuth was at best about 1 deg and at worst about 10 deg.

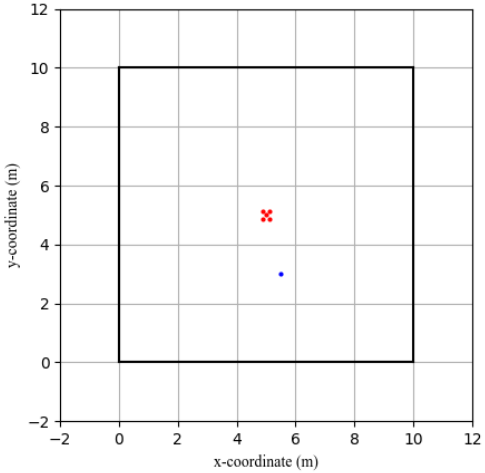


[3]

Rectangular Pyramid Array with TDOA

Based on the paper by Chen & Xu in 2019.

- Calculates the TDOA of all microphone-pairs using GCC-PHAT.
- Iteratively solves a model using Newton's method.
- Averages four estimates.

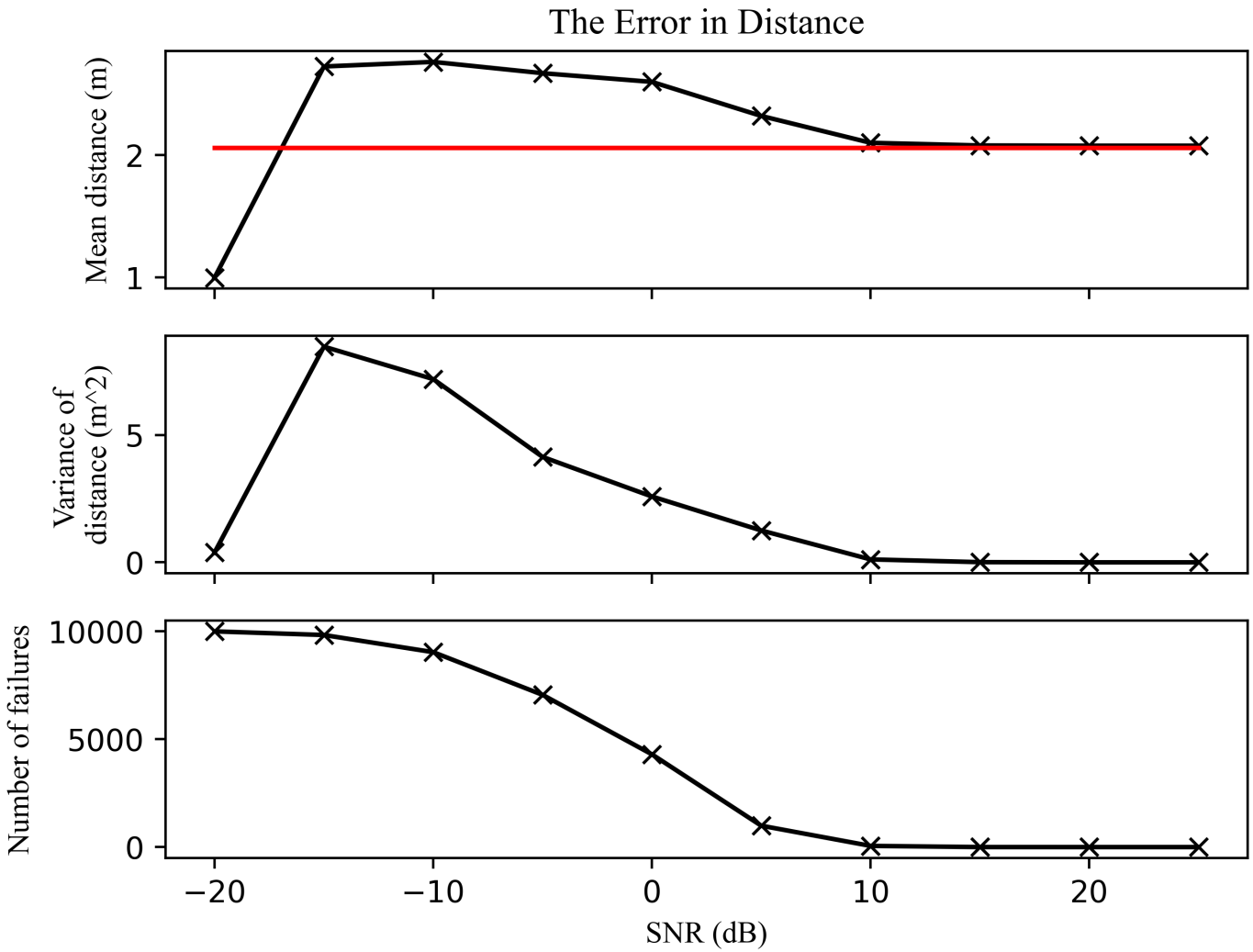


For each j-th microphone:

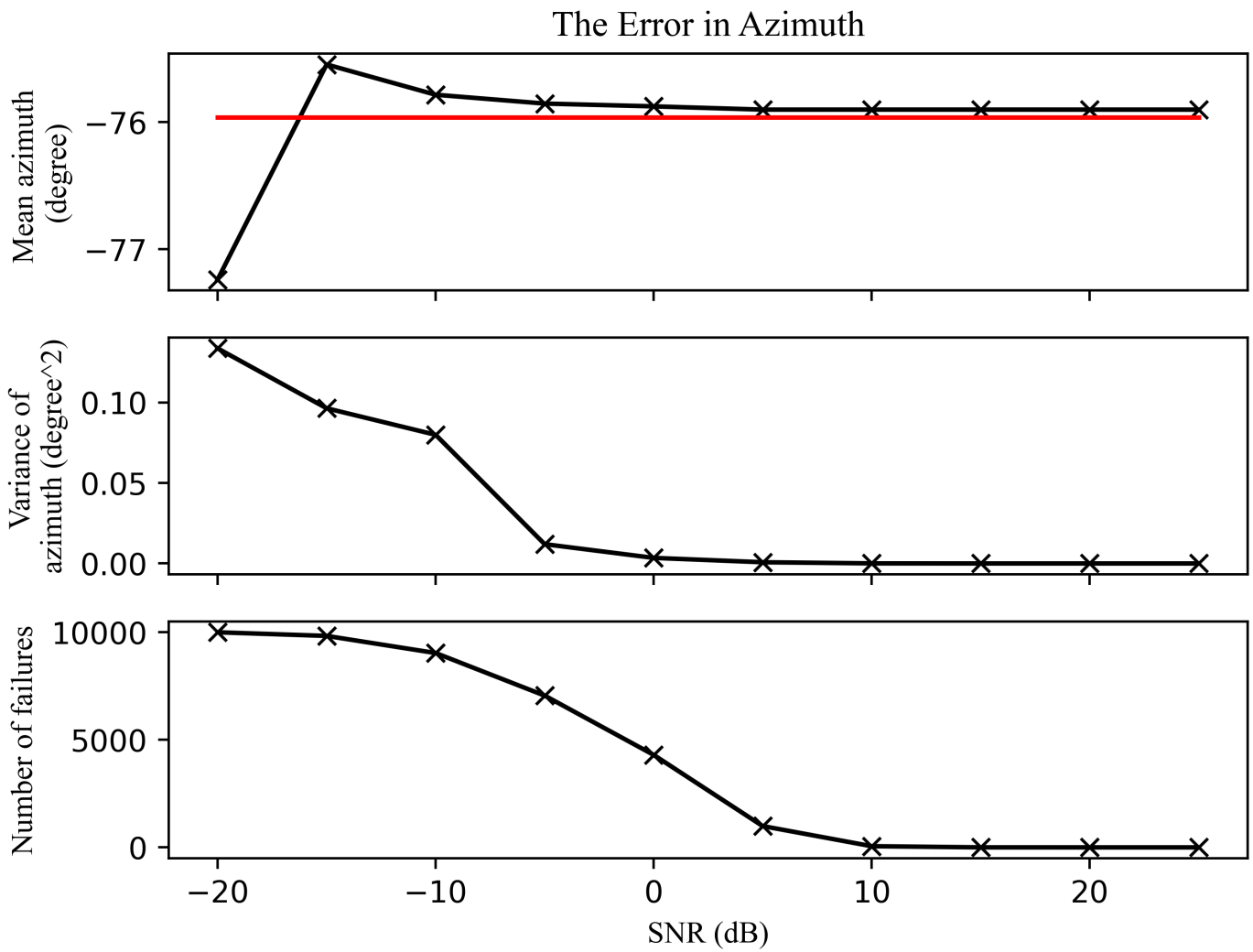
$$f_j(x,y,z) = \sqrt{(x-x_j)^2+(y-y_j)^2+(z-z_j)^2} - \sqrt{(x-x_0)^2+(y-y_0)^2+(z-z_0)^2} - c\Delta t_{j0} \tag{8}$$

Since it relies on an iterative algorithm, this method tends to fail to converge to a solution, especially for high noise.

Results for Distance



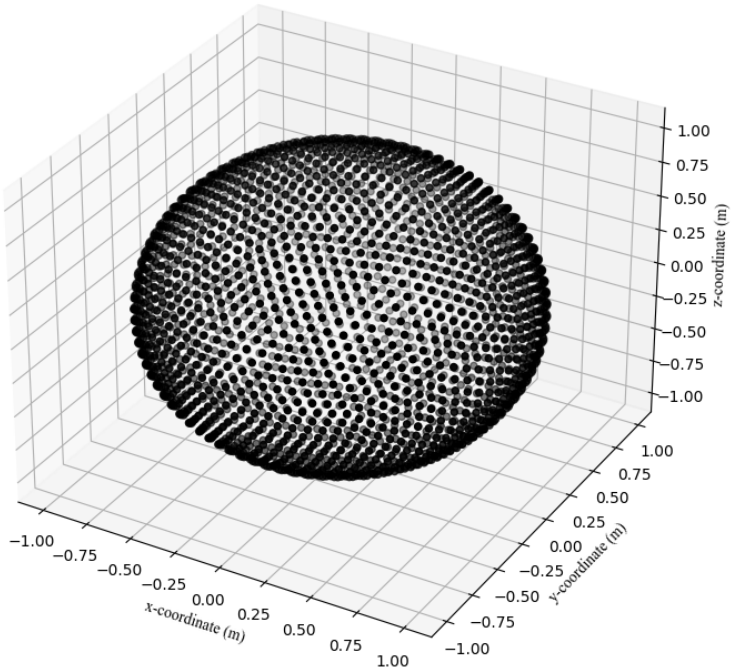
Results for Azimuth



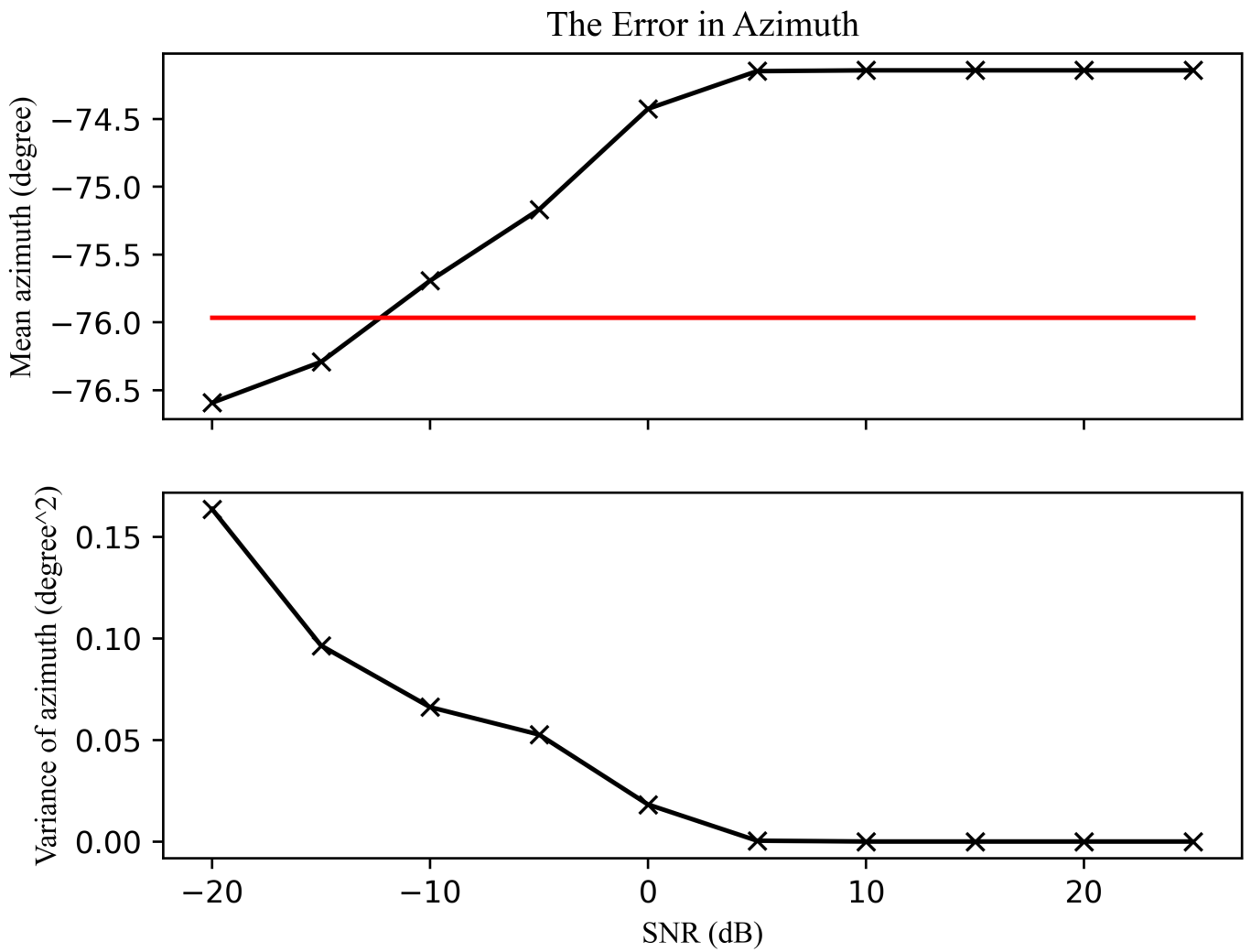
Beamforming with SRP-PHAT

Based on a few papers, mostly the one by Valin et al. in 2007.

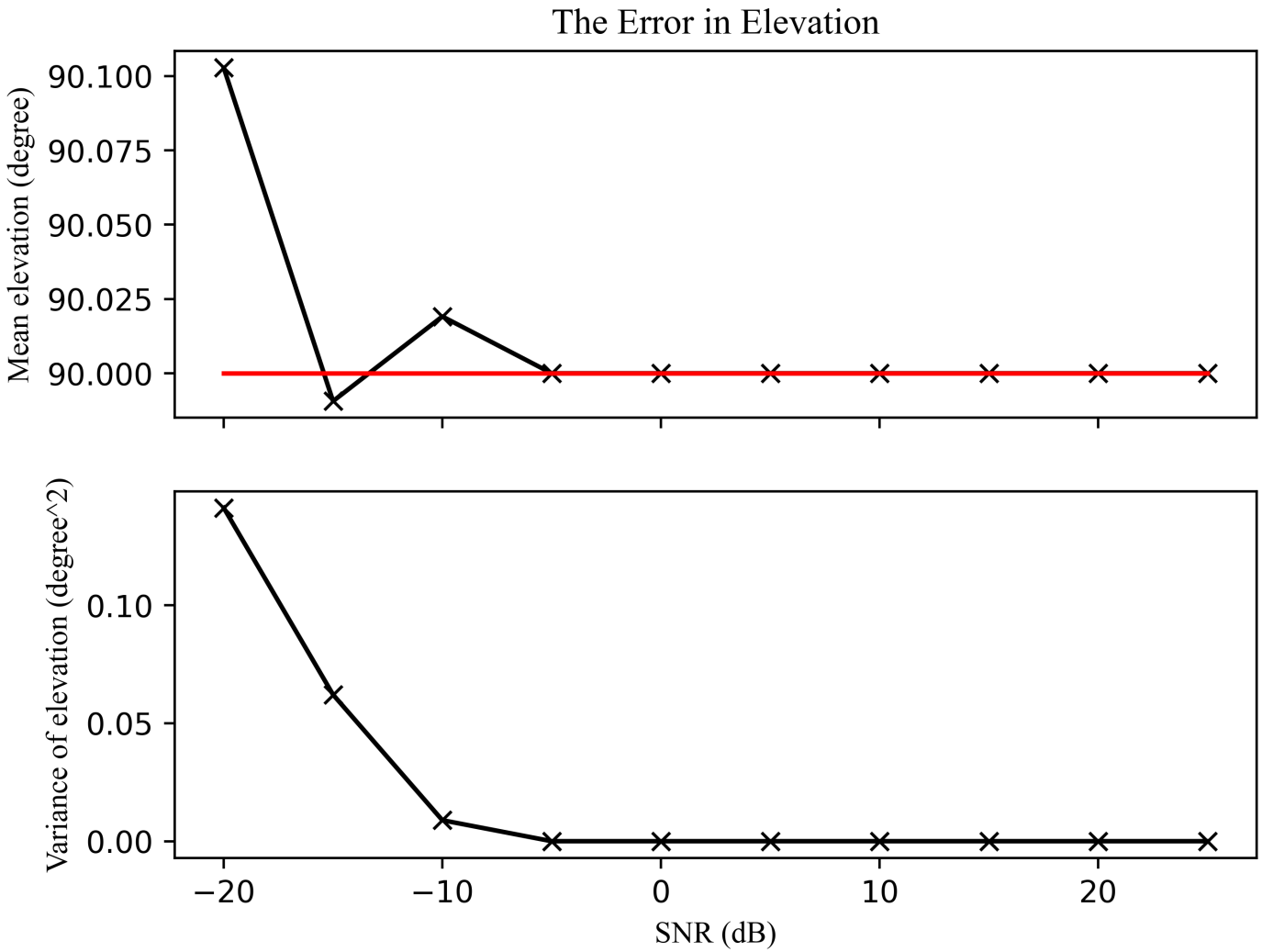
- The beamformer's energy is computed for every direction in a spherical search-grid of 2562 points.
- Estimates the direction as that corresponding to the highest energy in the energy-map.
- Only uses a coarse grid.



Results for Azimuth



Results for Elevation

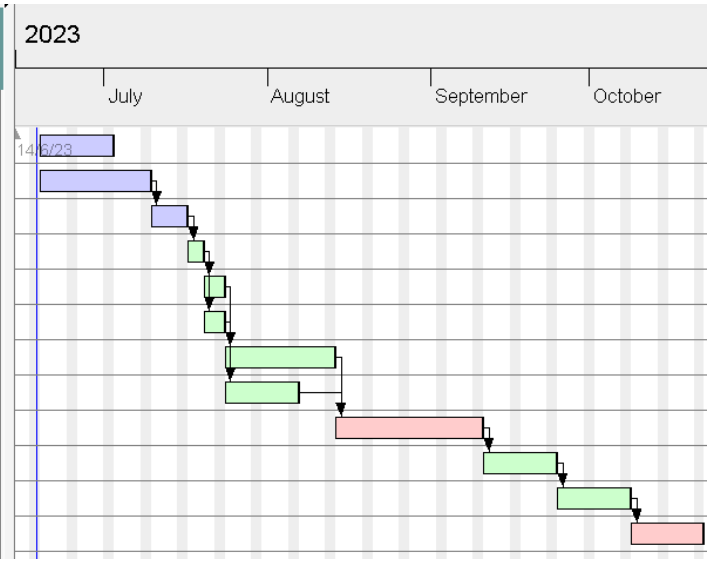


Progress So Far

- Defined the scope.
- Completed the literature-review.
- Investigated the three main methods.
- Did basic simulations of the main literature-examples.

Schedule

Name	Begin date	End date
Simulate examples for MUSIC	19/6/23	2/7/23
Investigate the best method.	19/6/23	9/7/23
Simulate a method in a realistic environment.	10/7/23	16/7/23
Design geometry.	17/7/23	19/7/23
Choose components.	20/7/23	23/7/23
Design electronics.	20/7/23	23/7/23
Write software.	24/7/23	13/8/23
Build the electronics.	24/7/23	6/8/23
Test in a real environment.	14/8/23	10/9/23
Modify the design.	11/9/23	24/9/23
Design the prototype on a NUgus.	25/9/23	8/10/23
Evaluate the final system.	9/10/23	22/10/23



Schedule

Winter:

- Simulate examples for MUSIC.
- Investigate the best method in terms of computation, noise, etc.
- Choose a method and simulate it in a realistic environment, e.g. a large hall.

Schedule

Next Semester:

- Design the geometry and the set-up of the array.
- Choose components, e.g. microcontroller, microphones, etc.
- Design the PCB and the physical set-up.
- Write software for the microcontroller/processor/etc.
- Build the array and the hardware, PCB, etc.
- Test the array in a real environment.
- Modify the design, software, etc.
- Set the design on a NUgus robot.
- Test and evaluate the final system in a robotics context.

References

S. Argentieri, P. Danès, and P. Souères.

A survey on sound source localization in robotics: From binaural to array processing methods.

Computer Speech & Language, 34(1):87–112, 2015.

Guoliang Chen and Yang Xu.

A Sound Source Localization Device Based on Rectangular Pyramid Structure for Mobile Robot.

Journal of Sensors, 2019:1–13, August 2019.

Keisuke Nakamura, Kazuhiro Nakadai, and Gökhan Ince.

Real-Time Super-Resolution Sound Source Localization for Robots.

In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 694–699, October 2012.

ISSN: 2153-0866.

Caleb Rascon and Ivan Meza.

Localization of sound sources in robotics: A review.

Robotics and Autonomous Systems, 96:184–210, 2017.

Jean-Marc Valin, François Michaud, and Jean Rouat.

Robust Localization and Tracking of Simultaneous Moving Sound Sources Using Beamforming and Particle Filtering.

Robotics and Autonomous Systems, 55:216–228, March 2007.