Master of Engineering Sciences

Intelligent Systems Engineering (ISI)

Sorbonne University

Advanced Audio Processing

Sound Source Localization with a Microphone Array: Beamforming Approaches

Students (Group 1):

Edouard David (28712992)
Lyes Mokhbi (21200513)

Academic Year 2024–2025

1 Introduction

This report details the implementation and analysis of sound source localization using beamforming approaches with an array of 8 omnidirectional MEMS microphones. In the previous lab session, sound propagation was characterized, and this information will now be used to determine the position of a sound source relative to a linear microphone array. We will work with a sampling frequency of $F_s = 20$ kHz and a buffer size of $BLK = 2048$.

2 Audio buffer acquisition

To start the process, the audio system must be started and an audio buffer captured. The resulting signals are then plotted versus time to visualize the captured data.

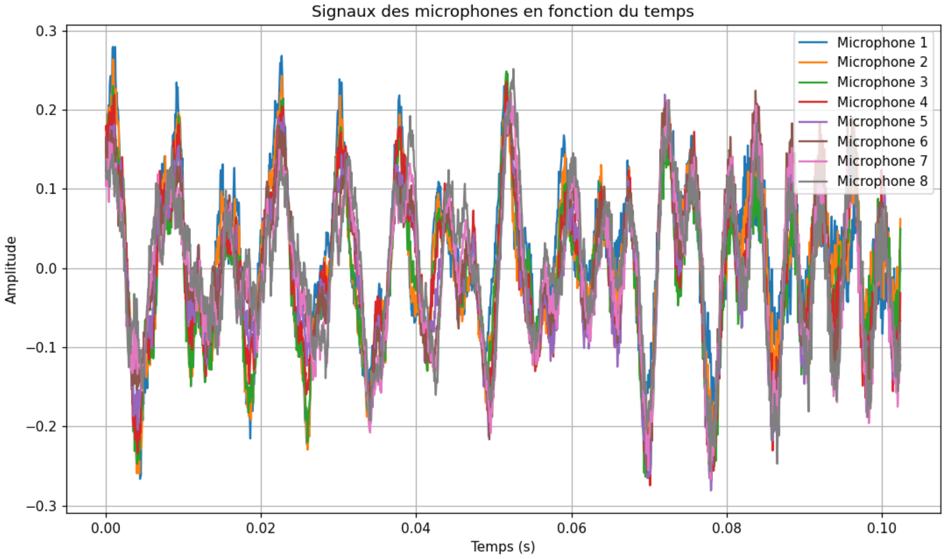


Figure 1: Time-domain signal of the audio buffer captured from the microphone array.

Comment: In Figure 1, we observe the time-domain signal of the captured audio buffer. This signal corresponds to the raw outputs of the microphones, which will later be processed using beamforming techniques.

3 Beamforming filters

Beamforming consists of applying a filter to each microphone signal and summing the filtered signals to form the beamformer output. The frequency response of the filter is given by:

$$W_n(f) = \mathcal{F}\{w_n(t)\} = e^{j2\pi \frac{f}{c} x_n \cos(\theta_0)} \tag{1}$$

where:

- $W_n(f)$ is the frequency response of the filter for the n-th microphone,
- f is the frequency,

- c is the speed of sound in air,
- x_n is the position of the n-th microphone,
- θ_0 is the steering (focus) angle of the beamformer.

3.1 Microphone positions

The position x_n of the microphones can be expressed as a function of the microphone index n and the spacing d between microphones. The first microphone is numbered 0, and the origin is placed at the center of the array. The position is given by:

$$x_n = \left(n - \frac{N}{2}\right) \cdot d \tag{2}$$

3.2 Beamformer filter function

To compute the frequency response of the filter for each microphone, we define a Python function `beam_filter`. This function takes as input the microphone array, a frequency vector, the steering angle θ_0, and the microphone index n, and returns the corresponding frequency response.

The implementation of this function can be found in the Jupyter notebook attached to this report.

3.3 Comparison of frequency responses for two microphones at $\theta_0 = 0°$

The effect of the steering angle θ_0 on the filter responses is studied by comparing the filters associated with two different microphones when $\theta_0 = 0°$, for frequencies ranging from 0 to 5 kHz. The results are presented below.

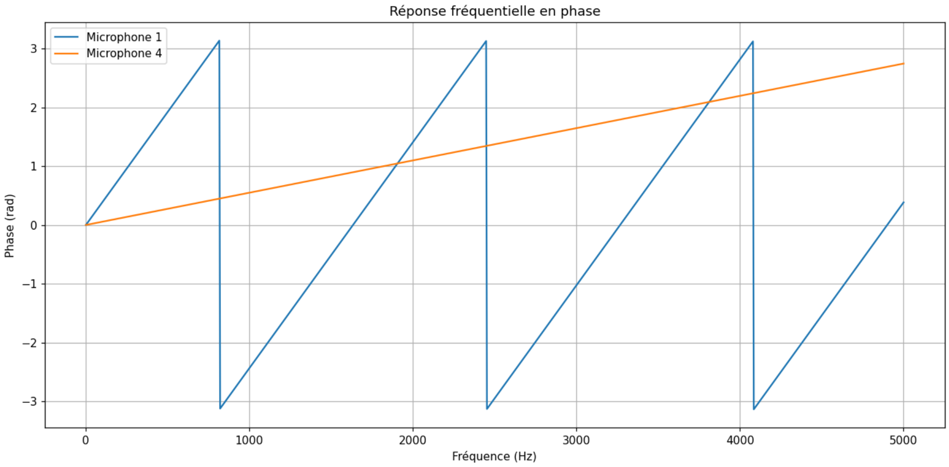


Figure 2: Frequency response of the filters applied to Microphone 1 (blue) and Microphone 4 (orange).

Comment: In Figure 2, the phase curve of Microphone 4 (orange) shows a linear progression without apparent discontinuities. This corresponds to a typical phase response *of a filter applied to a microphone located at a specific position in space. This linear phase indicates that the signal is processed uniformly in terms of delay, which is often the case for a microphone placed at a particular angle relative to the sound source. The phase of Microphone 1 (blue) shows a linear progression but includes discontinuities at regular intervals. These discontinuities correspond to phase "wrap" jumps from $-\pi$ to π, which is a normal property of phase angle representations. This may be due to complex interaction between the filter and how the signal is perceived by this microphone, or to a digital signal processing effect.*

3.4 Comparison of filters obtained for $\theta_0 = 90°$

At $\theta_0 = 90°$, the sound source is positioned perpendicular to the array axis. In this configuration, all microphones theoretically receive the signal without propagation time differences. This means there is no relative delay between microphones, and therefore the filters apply little or no phase correction.

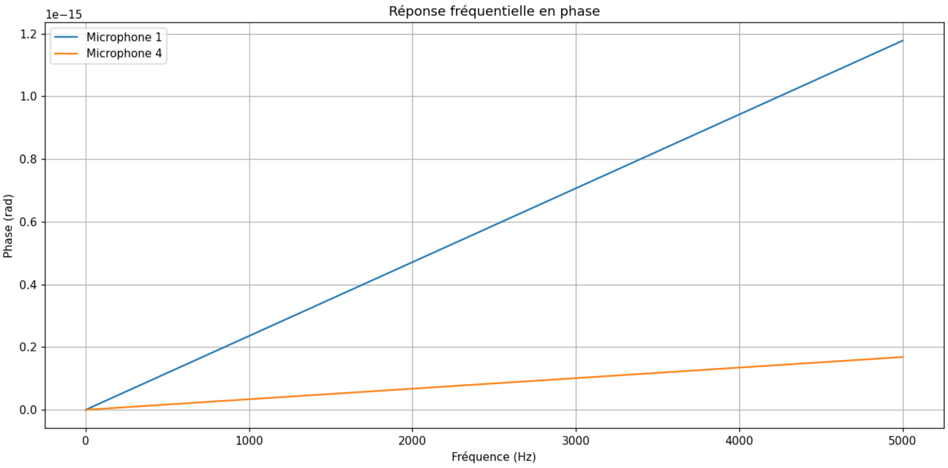


Figure 3: Frequency response for $\theta_0 = 90°$.

Comment: In Figure 3, at $\theta_0 = 90°$, the phase of the filters is almost zero, indicating that there is no need for significant correction. Since all microphones are symmetrically positioned with respect to the source, the signal is perceived similarly by each microphone, and the phase adjustments are minimal.

4 Beamforming algorithm

The beamforming process consists of several steps to localize a sound source at a specific frequency. These steps include acquiring an audio frame, performing an FFT, and computing the beamformer outputs for different steering directions.

4.1 Audio buffer acquisition and FFT computation

The first step is to acquire an audio buffer and compute its FFT. The FFT of the buffer reveals the frequency components of the signals from all microphones. The Fourier transform is given by:

$$M_{\text{fft}}(k) = \sum_{t=0}^{BLK-1} m(t) \cdot e^{-j2\pi kt/BLK} \tag{3}$$

where $M_{\text{fft}}(k)$ is the frequency component at index k, $m(t)$ is the microphone signal at time t, and BLK is the buffer size.

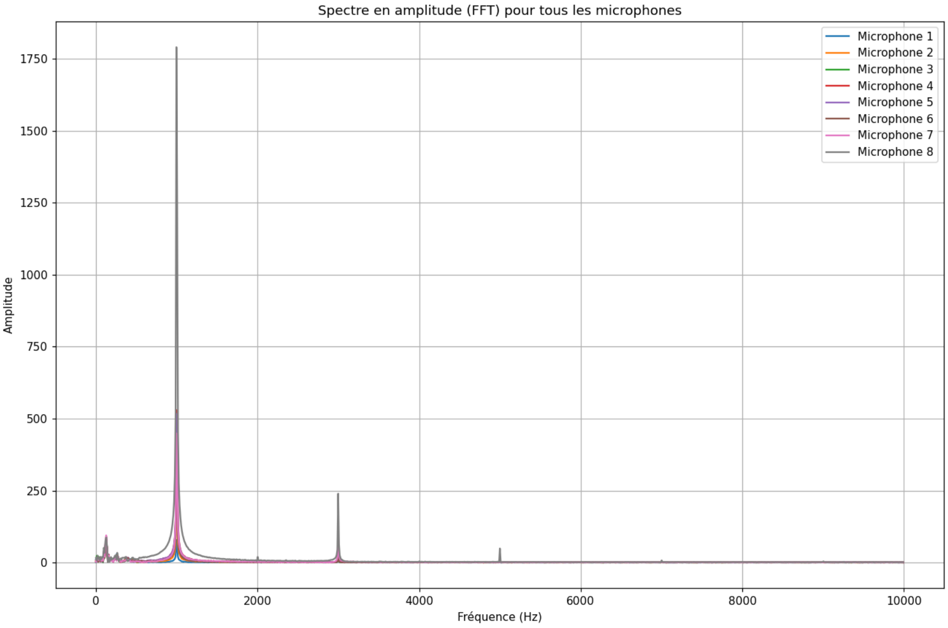


Figure 4: FFT of an audio buffer while emitting a 1 kHz sinusoidal tone.

Comment: In Figure 4, a clear and dominant peak is visible at frequency $F_0 = 1$ kHz, as expected, since a pure sinusoidal signal at this frequency was emitted. This confirms that the signal was captured by all microphones. The amplitudes of the different microphones are very close at the main frequency (1 kHz). We also observe peaks at multiples of F_0 (2 kHz, 3 kHz); these peaks are likely due to noise present in the room.

4.2 Frequency selection and filter application

Once the frequency components are obtained after applying the Fourier transform, it is necessary to select the frequency of interest, here $F_0 = 1\,\text{kHz}$, and extract the FFT values associated with this frequency for each microphone. This step relies on identifying the index k_0 corresponding to the closest frequency to F_0 in the positive-frequency vector obtained during the FFT.

The index k_0 is determined by finding the position of the peak closest to F_0 in the positive-frequency vector positive_freqs, using:

$$k_0 = \arg\min_k |\text{positive\_freqs}[k] - F_0| \tag{4}$$

Once this index k_0 is identified, the exact frequency associated with that index is obtained as:

$$\text{exact\_freq} = \text{positive\_freqs}[k_0] \tag{5}$$

From this exact frequency, the FFT values associated with all microphones at that frequency are extracted. These values are represented by the vector M, which contains the amplitudes and phases of the signals from each microphone for frequency F_0. In notation:

$$M[k_0] = [M_1[k_0], M_2[k_0], \ldots, M_N[k_0]] \tag{6}$$

where each $M_n[k_0]$ represents the FFT component for the n-th microphone at frequency F_0.

The results obtained for $k_0 = 102$, corresponding to an exact frequency of 996.09 Hz, are as follows:

FFT amplitudes and phases for each microphone

Microphone	**FFT Amplitude**	**FFT Phase**
1	78.62042	−2.0243607
2	239.8363	−1.3679484
3	419.63046	−1.113776
4	529.75995	−0.9095199
5	516.78784	−0.5971969
6	338.89374	0.00567186
7	448.90628	1.723821
8	1790.6444	2.604943

Interpretations:

Index $k_0 = 102$ and exact frequency:

- Index $k_0 = 102$ corresponds to a frequency of 996.09375 Hz, which is very close to the expected source frequency (1 kHz).
- This small difference is due to the frequency resolution limited by the FFT buffer size (BLK) and the sampling frequency (Fs).

Microphone amplitudes:

- FFT amplitudes vary across microphones, ranging from 78.62 (Microphone 1) to 1790.64 (Microphone 8). This may reflect several factors:
- Microphones closer to the source may capture a higher-amplitude signal.
- Amplitudes vary depending on the orientation of the array relative to the sound source.

Microphone phases:

- Phases vary across microphones, ranging from -2.02 rad (Microphone 1) to 2.60 rad (Microphone 8).
- Each microphone captures the signal at a slightly different time, resulting in a phase shift.

Additional observations:

- The amplitude is proportional to the distance to the source. Microphones closer to the source capture higher amplitudes.
- The phase is directly related to the propagation delay of the sound signal reaching each microphone (its position).

4.3 Computation and application of filters for each microphone

Once the frequency of interest f_0 and the main direction θ_0 are defined, it is necessary to compute the filters associated with each microphone in the array. These filters are obtained from the function `beam_filter`, which, for each microphone, computes a filtering factor based on the position θ_0 and the frequency f_0 obtained previously.

The filters are then applied to the FFT component of each microphone, yielding a vector M_{filtered} containing the results after filtering.

The filter computation (see the Jupyter notebook attached to the report for more details) for the n-th microphone is done as follows:

$$\text{filter\_value} = \text{beam\_filter}(\text{antenne}, [f_0], \theta_0 = 0, \text{mic\_nb} = n)[0]$$

Then the filter is applied to each microphone FFT value $M[n]$:

$$M_{\text{filtered}}[n] = M[n] \times \text{filter\_value}$$

The results after applying the filters to each microphone show the filtered contributions of each microphone signal, whose phase and amplitude are adjusted according to the applied filter.

Example of results after filter application:

Microphone	**Components of M_{filtered}**
1	$-18.43344065 + 76.42891507j$
2	$48.05455807 + 234.9727802j$
3	$362.39204926 + 211.56965028j$
4	$495.40447064 - 187.66990246j$
5	$213.64548864 - 470.55848847j$
6	$-22.26316136 - 338.16166154j$
7	$237.54011986 - 380.90880869j$
8	$603.76639869 - 1685.78564487j$

Interpretations:
Vector M_{filtered} after filtering:

- The vector M_{filtered} contains the microphone contributions after filtering, expressed in complex form.
- Each complex element has two parts: magnitude (amplitude) and phase.

Microphone amplitude and phase:

- Microphone amplitudes vary significantly, which may be due to factors such as microphone distance to the sound source or acoustic reflections in the environment. **(see Figure 5 for a histogram of amplitudes)**.

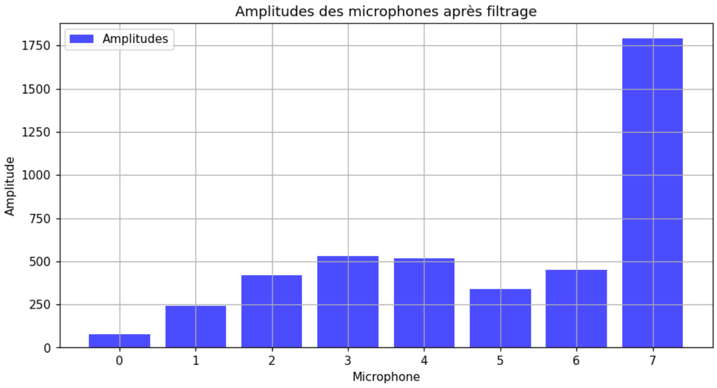


Figure 5: Histogram of microphone amplitudes.

- Phases also vary from one microphone to another, reflecting the phase shift due to each microphone's relative position to the sound source. **(see Figure 6 for phase evolution)**.

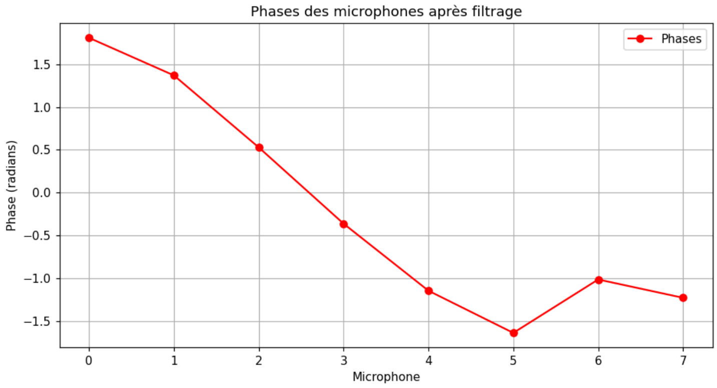


Figure 6: Phase evolution of the microphones.

4.4 Power $P(\theta_0)$ computation for different directions

In this section, we computed the power $P(\theta_0)$ for a source emitting in two different directions: one close to $\theta_0 = 0°$ and another farther away at $\theta_0 = 45°$. The results are presented in the table below.

$$P(\theta_0) = \left| \sum_{n=0}^{N-1} W_n(f) \cdot x_n \right|^2 \tag{7}$$

where $P(\theta_0)$ is the beamformer output power for direction θ_0, and $W_n(f)$ is the frequency response of the filter applied to the n-th microphone.

Direction θ_0	**Power $P(\theta_0)$**
Near ($\theta_0 = 0°$)	10,138,983.77
Far ($\theta_0 = 45°$)	15,500,366.36

Table 1: Power $P(\theta_0)$ for different source directions

Interpretations:

- Contrary to what one might expect, the power for the far direction ($\theta_0 = 45°$) is higher than for the near direction ($\theta_0 = 0°$).

Possible causes:

- **Acoustic reflections or interference:** Reflections in the environment could introduce significant contributions from a direction different from $\theta_0 = 0°$, such as $\theta_0 = 45°$, increasing the observed power.
- **Incorrect estimation of the true source direction:** If the sound source is actually closer to $\theta_0 = 45°$, phases would be better aligned in that direction, explaining the higher power.
- **Filter or microphone calibration:** Imperfect calibration of microphones or filters could cause phase alignment errors, affecting expected results and producing a higher power in an off-target direction.
- **Noise or parasitic sources:** Noise or interference coming from other directions could unexpectedly contribute to the measured power in the far direction ($\theta_0 = 45°$).

Additional observations:

- The observed power depends on how signals are filtered and on the phase shift between microphones.
- External factors such as reflections or calibration errors can significantly influence the results and must be considered in the analysis.

4.5 Power computation and directional localization

For each direction θ_0, the power of the beamformer output is computed and plotted as a function of θ_0 to detect the direction of maximum power. The power is given by Equation 7.

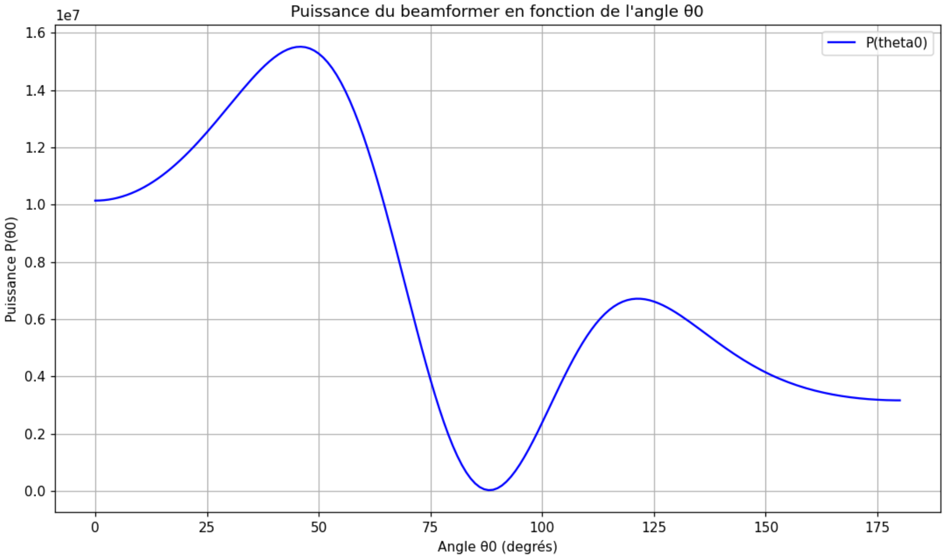


Figure 7: Beamformer output power as a function of θ_0.

Interpretation of Figure 7:

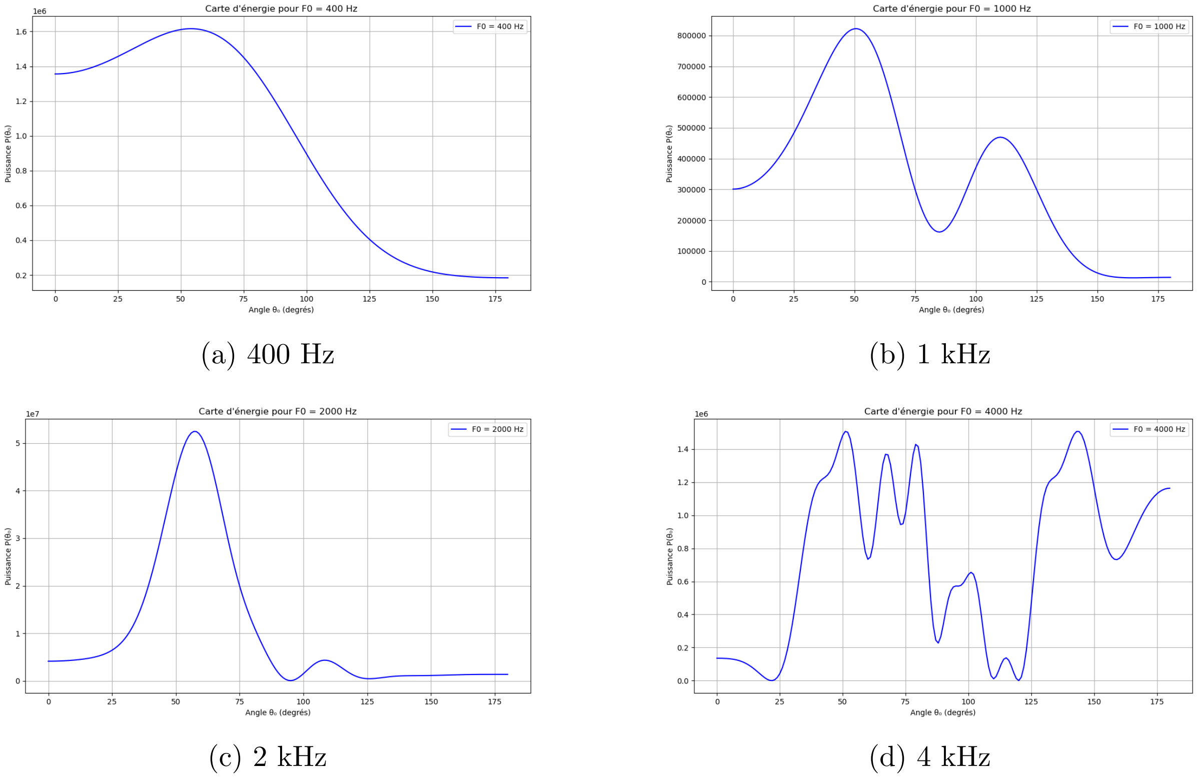
- A power peak is clearly visible around $\theta_0 = 45°$.
- A decrease in power is observed for angles far from this direction, with a minimum around $\theta_0 = 90°$.
- This suggests that the sound source is located around $\theta_0 = 45°$.
- In a non-anechoic environment, reflections from walls or surrounding objects can create virtual sources. These reflected sources appear to come from a different direction, for example 45°.
- If reflections are strong, the beamformer may interpret them as the main source.

5 Beamformer performance analysis

In this section, we analyze beamformer performance by examining energy maps for different source frequencies and by evaluating the accuracy of position estimation in the case of a moving source.

5.1 Energy maps for fixed frequencies

We generated energy maps for fixed frequencies $F_0 = 400\,\text{Hz}, 1\,\text{kHz}, 2\,\text{kHz}, \text{and } 4\,\text{kHz}$, emitted from a fixed and arbitrary position. These maps make it possible to visualize the relative signal intensity as a function of position in the listening field.



(a) 400 Hz

(b) 1 kHz

(c) 2 kHz

(d) 4 kHz

Figure 8: Energy maps for $F_0 = 400\,\text{Hz}, 1\,\text{kHz}, 2\,\text{kHz}, \text{and } 4\,\text{kHz}$ for a source emitting at an angle of $\theta_0 = 0°$.

Interpretation:

- For frequency $F_0 = 400\,\text{Hz}$ (see Figure 8a), the curve shows a maximum intensity around $\theta_0 = 50°$, which is consistent with expectations, given that the beamformer is steered toward $\theta_0 = 0°$. This offset can be attributed to slight imprecision in the phone orientation. The intensity decreases progressively as θ_0 moves away from 50°, indicating reduced sensitivity outside the main axis. The attenuation outside $\theta_0 = 50°$ is not very abrupt, which is consistent with low directivity at this frequency.

- For frequency $F_0 = 1\,\text{kHz}$ (see Figure 8b), the maximum intensity is also observed near $\theta_0 = 50°$, but attenuation is more pronounced as we move away from that direction. In addition, a second lobe appears around $\theta_0 = 120°$. This phenomenon may be due to imprecision in source localization, probably because the emitted wave frequency is above the system's limit frequency. Indeed, the system is frequency-limited and operates only below f_{lim}. This behavior is due to the microphone spacing becoming too large compared to the wavelength λ, resulting in multiple periods being captured by the first microphone before the wave reaches the second. It then becomes impossible to correctly determine the speed of sound based on extrema or zero-crossings. The limit frequency f_{lim} can be computed from the setup characteristics with the following condition:

  $$\lambda > d$$

  Since $\lambda = \frac{c}{f}$, the frequency must satisfy:

  $$f < \frac{c}{d}$$

  In our case, $f_{lim} = 810\,\text{Hz}$.

- For frequency $F_0 = 2\,\text{kHz}$ (see Figure 8c), the intensity continues to decrease as θ_0 moves away from 0°, but with a sharper drop. We observe two lobes similar to the previous case, but the main lobe is narrower and centered around $\theta_0 = 60°$. Moreover, the amplitude of the second lobe is weaker compared to that observed for $F_0 = 1\,\text{kHz}$. This indicates improved directivity, as expected.

- Finally, for frequency $F_0 = 4\,\text{kHz}$ (see Figure 8d), we observe oscillations characteristic of a loss of directivity, which is expected at high frequencies due to spatial aliasing.

- In summary, as frequency increases, beam directivity improves, with faster attenuation outside the main direction. For a high frequency (4 kHz), directivity becomes sharper, but the microphone array still detects the source in all directions.

5.2 Position estimation for a moving source

The estimation method is based on computing the maximum power $P_{\max}$ from the signals received by multiple microphones arranged in an array. The procedure is described mathematically as follows.

Let **audio** be a matrix representing the received signals, where each column corresponds to one microphone. The first step is to compute the dominant frequency index k_0 for each signal:

$$k_0 = \arg\max\left(\left|\mathcal{F}\left(\mathbf{audio}(:,1)\right)\right|\right)$$

where $\mathcal{F}$ denotes the discrete Fourier transform (FFT). Next, for each angle j from 0 to 360, we compute the summed power associated with each direction:

$$P_j = \left|\sum_{i=1}^{8}\left(\text{beam\_filter}(\mathbf{antenne}, f_{k_0}, j, i) \cdot \mathcal{F}(\mathbf{audio}(:,i))_{k_0}\right)\right|^2$$

Here, $\text{beam\_filter}(\mathbf{antenne}, f_{k_0}, j, i)$ is a beam filter that depends on the array, the frequency f_{k_0}, and the angle j, and $\mathcal{F}(\mathbf{audio}(:,i))_{k_0}$ is the FFT value at the dominant frequency k_0 for the i-th microphone.

Finally, the estimated source position corresponds to the angle $j_{\max}$ where the power is maximal:

$$P_{\max} = \arg\max(P)$$

This makes it possible to determine the direction of the moving source at each time instant.

For a fixed frequency $F_0 = 1\,\text{kHz}$, we moved a source around the microphone. The estimated position as a function of time is shown in the following figure.

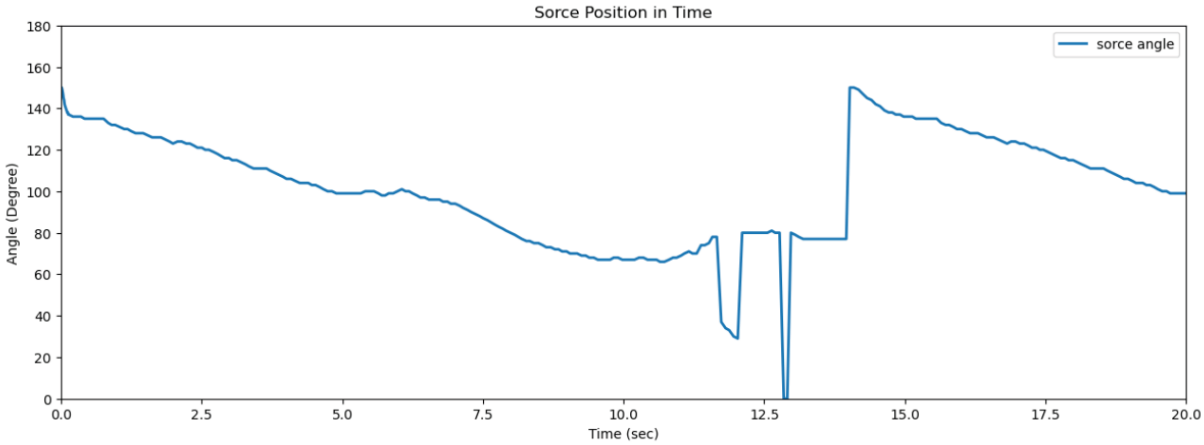


Figure 9: Estimated position of the moving source ($F_0 = 1\,\text{kHz}$) as a function of time.

Interpretation: Overall, the estimation method seems robust, with continuous angle tracking over most of the observed period. The moving source primarily travels along a circular path, which results in progressive variations of the angle over time.

However, around $t = 12.5\,\text{s}$, a rapid drop in the angle is followed by a sudden increase. After these disturbances, the curve resumes a regular decreasing trend, which continues until the end of the 20-second interval.

This anomaly can be explained by the use of **buffers** when retrieving the signal. Indeed, closer inspection of the curve in the interval $[15\,\mathrm{s}; 20\,\mathrm{s}]$ shows that its evolution is identical to that observed in the initial interval $[0\,\mathrm{s}; 5\,\mathrm{s}]$. This suggests that the last part of the signal corresponds to the beginning of the signal, displayed again.

Conclusion

This report explored the performance of a microphone-array and beamforming system for sound source localization. The results show that amplitude and phase variations across microphones are influenced by the source position, acoustic reflections, and equipment calibration. The directional power analysis revealed unexpected results, suggesting the impact of reflections and source estimation errors. Finally, improved directivity with increasing frequency was observed, although artifacts related to system limitations appeared at higher frequencies.