# Direction of Arrival Estimation with a Circularly Moving Microphone

Jeremy Lawrence[1], Jens Ahrens[2] and Nils Peters[1]

[1] *International Audio Laboratories Erlangen*, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany*

[2] *Division of Applied Acoustics, Chalmers University of Technology, Gothenburg, Sweden*

*Email: jeremy.lawrence@fau.de*

## Introduction

Estimating the direction of arrival (DOA) of sound sources is an important task of sound field analysis which is conventionally performed using two or more stationary microphones. In recent years, various single-microphone localization approaches have been proposed to reduce the cost and memory needs of multi-microphone setups. Many of these utilize a scattering body with known DOA-dependent scattering characteristics and require prior knowledge of the sound sources to be localized [1, 2, 3].

The introduction of microphone movement allows for DOA estimation without these requirements. This was demonstrated in [4], where a signal was constructed from a circular microphone array using circular sampling, i.e., taking the first sample from one microphone and the following samples from the neighboring microphones in a circular fashion. Although this approach does not utilize movement, the resulting signal is equivalent to that captured by a circularly moving microphone. This artificial motion introduces periodic shifts in frequency due to the Doppler effect. The DOA of sound sources can subsequently be estimated by estimating the phase of these periodic shifts. In [5] this approach was verified in practice using a circularly moving microphone rotating at a maximum speed of approximately 17 rotations per second. However, localization was only performed using known, single-frequency signals.

The DOA estimation approach proposed in this paper not only allows for the localization of unknown wideband and single-frequency sources but also enables the reconstruction of the original signal. The algorithm was verified in practice in a low reverberant environment using the rotating equatorial microphone (REM) depicted in Fig. 1. In our experiments, rotational speeds between 24, 34, and 42 rotations per second were employed. More details regarding the REM prototype can be found in [6]. Please note that this paper corresponds to a condensed version of the article found in [7]. For a more comprehensive description of the algorithm and the results presented herein, please refer to the full article.

## Problem formulation

Consider the setup depicted in Fig. 2, where a circularly moving microphone is situated in a free field which contains one sound source emitting a single frequency $f_{\mathrm{src}}$. The source is placed at azimuth $\varphi$ relative to the initial microphone position and at a sufficient distance

**Figure 1:** Photograph of the REM prototype.

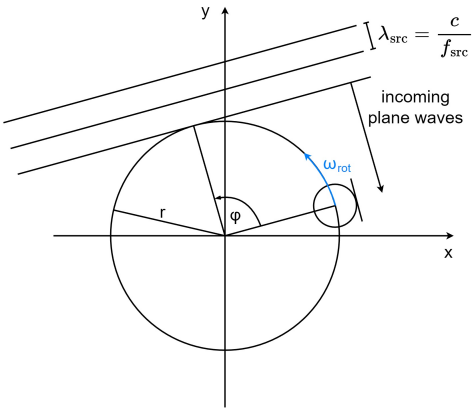



**Figure 2:** Rotating microphone in a sound field composed of monochromatic plane waves of frequency $f_{\mathrm{src}}$ arriving at azimuth $\varphi$ relative to the initial microphone position.

such that all incoming sound waves can be assumed to be plane waves with a constant amplitude. The radius and the angular velocity of the rotation are selected as $r$ and $\omega_{\mathrm{rot}} = 2\pi f_{\mathrm{rot}}$. Under these conditions, the frequency observed by the microphone is sinusoidally shifted due to the Doppler effect and can be expressed as

$$f_{\mathrm{obs}}(t, f_{\mathrm{src}}, \varphi) = \left(1 - \frac{2\pi r f_{\mathrm{rot}} \cdot \sin(2\pi f_{\mathrm{rot}} t - \varphi)}{c}\right) \cdot f_{\mathrm{src}}\,, \tag{1}$$

where $c$ denotes the speed of sound. The signal captured by the microphone therefore corresponds to a sinusoidally frequency modulated waveform with carrier frequency $f_{\mathrm{src}}$ and modulation frequency $f_{\mathrm{rot}}$. Mathe-

*The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg and Fraunhofer IIS.

matically, this can be expressed as

$$x(t, f_{\mathrm{src}}, \varphi) = A_0 \cos(2\pi f_{\mathrm{src}} t + \beta \cos(2\pi f_{\mathrm{rot}} t - \varphi) + \phi_0)\,, \quad (2)$$

where $A_0$ represents the amplitude of the plane waves with arbitrary initial phase $\phi_0$ and $\beta = 2\pi r \cdot f_{\mathrm{src}}/c$ denotes the modulation index, which quantifies the degree to which the frequency is being shifted.

The direction-dependent frequency response of the microphone further impacts the signal that is being captured. For the sake of simplicity, we neglect the phase response of the microphone in this paper and only consider the direction-dependent magnitude response, which we denote as $|H(f_m, \varphi_m)|$. Here, $f_m$ denotes the frequency of interest and $\varphi_m$ refers to the DOA of a sound source relative to the front of the microphone. For example, the on-axis magnitude response of the microphone is given by $|H(f_m, 0)|$. When taking into consideration the microphone's direction-dependent magnitude response, we find that the amplitude captured by the microphone changes periodically. This can be modeled by replacing $A_0$ from (2) with the following term:

$$A_0(t, f_{\mathrm{src}}, \varphi) = A_0 \cdot H(f_{\mathrm{obs}}(t, f_{\mathrm{src}}, \varphi), \varphi - 2\pi f_{\mathrm{rot}} t)\,, \quad (3)$$

where $f_{\mathrm{obs}}(t, f_{\mathrm{src}}, \varphi)$ is given by (1). The signal captured by the microphone therefore corresponds to a frequency and amplitude-modulated version of the original signal. Both modulation terms are periodic with a period of $f_{\mathrm{rot}}$ and an initial phase $\varphi$. Direct estimation of the phase of the modulation is only possible for monochromatic and known signals, therefore we require a more elaborate approach to localize unknown and arbitrary signals. The idea we propose is to compensate for the distortions introduced by the microphone rotation. More specifically, we will compensate the distortions for multiple DOA guesses and select the DOA guess whose associated compensated signal features the least distortion. This, however, introduces two new challenges, namely how to perform DOA-dependent distortion compensation and how to quantify the distortion present in each compensated signal without prior knowledge of the source signal. These two points will be addressed in the remaining sections.

## Modulation compensation

Since the distortions introduced by the microphone movement are forms of amplitude and frequency modulation, we will refer to the compensation of these effects as amplitude and frequency *unmodulation* for a given DOA.

We denote the discrete signal captured by the microphone as $x(nT)$, where $T$ is the sampling interval and $n \in \{0, 1, ..., N-1\}$ is the $n$-th sample of the $N$-length signal. We further define $X(k)$ as the discrete Fourier transform (DFT) of $x(nT)$, where $k \in \{0, 1, ..., N/2\}$ is the $k$-th frequency bin. To perform amplitude and frequency unmodulation, we define the following transform:

$$y(nT, \varphi) = \frac{1}{N} \sum_{k=0}^{N/2} X(k) \cdot e^{j2\pi \frac{k}{N} n} \cdot z(n, k, \varphi)\,, \quad (4)$$

where $y(nT, \varphi)$ is the transformed signal for a given DOA $\varphi$ and $z(n, k, \varphi)$ is a transformation function, which we will define in the following. If we set $z(n, k, \varphi) = 1$ then (4) corresponds to the inverse DFT of $X(k)$, resulting in $y(nT, \varphi) = x(nT)$. By choosing $z(n, k, \varphi)$ appropriately, we can perform a *modulated* inverse DFT which uses modulated complex sinusoids as opposed to complex sinusoids as its basis functions. This allows us to amplitude and frequency modulate arbitrary signals, as opposed to single-frequency signals discussed in the previous section, using arbitrarily shaped and frequency-dependent modulation functions. By choosing the correct modulation shapes, the modulation introduced by the microphone rotation can be canceled out for a given DOA, resulting in unmodulation.

## Amplitude unmodulation

To obtain the correct $z(n, k, \varphi)$ for amplitude unmodulation, it is beneficial to first address how $z(n, k, \varphi)$ would have to be chosen to model the frequency-dependent amplitude modulation introduced by the microphone rotation. Eq. (3) shows how a single frequency is amplitude modulated. Therefore, we can derive the amplitude-modulated basis functions by applying (3) to each DFT basis function at its corresponding center frequency. This is performed by setting $z(n, k, \varphi)$ as

$$z(n, k, \varphi) = A_0(nT, \frac{k}{NT}, \varphi)\,. \quad (5)$$

Performing amplitude unmodulation requires selecting $z(n, k, \varphi)$ such that the amplitude modulation introduced by (5) cancels out. This is easily achieved by selecting

$$z(n, k, \varphi) = \frac{1}{A_0(nT, \frac{k}{NT}, \varphi)}\,. \quad (6)$$

## Frequency unmodulation

Compensating for the frequency modulation introduced by the microphone rotation is more cumbersome since it affects the signal in a more complex fashion than amplitude modulation, which only affects the envelope of each DFT basis function. To derive the necessary $z(n, k, \varphi)$ for frequency-dependent frequency unmodulation, it is once again beneficial to first simulate the frequency modulation introduced by the microphone rotation. In light of this, we reformulate (2) as

$$x(t, f_{\mathrm{src}}, \varphi) = A_0 \cos(2\pi f_{\mathrm{src}}(t + t_{\mathrm{shift}}(t, \varphi)) + \phi_0)\,, \quad (7)$$

where $t_{\mathrm{shift}}(t, \varphi) = r/c \cdot \cos(2\pi f_{\mathrm{rot}} t - \varphi)$. This reformulation makes it clear that a sampled modulated single-frequency signal is identical to an unmodulated signal sampled at times $nT + t_{\mathrm{shift}}(nT, \varphi)$. By formulating these time shifts as a complex exponential in a similar manner to the complex DFT basis functions, the frequency-dependent frequency modulation introduced by the microphone rotation can be modeled by setting

$$z(n, k, \varphi) = e^{j2\pi \frac{k}{N} \cdot \frac{t_{\mathrm{shift}}(nT, \varphi)}{T}}\,. \quad (8)$$

To perform frequency unmodulation we require time shifts which cancel out the time shifts from (8). To our

knowledge, these time shifts cannot be computed analytically, therefore they are instead obtained using the following optimization:

$$t_{\text{shift}}(nT, \varphi) = \underset{x}{\operatorname{argmin}} \left| r \cos(2\pi f_{rot}\,(nT + x) - \varphi) + c\,x \right| . \tag{9}$$

A detailed derivation of this formula can be found in [7].

These insights now allow us to amplitude and frequency unmodulate a signal for a given DOA guess by applying amplitude unmodulation (Eqs. (4) and (6)) and subsequently frequency unmodulation (Eqs. (4), (8) and (9)).

## Quantifying amplitude modulation

Let us now cover how the presence of amplitude modulation in the unmodulated signals is quantified. Since the microphone employs rotational speeds between 24 to 42 rotations per second, the majority of real-world signals can be assumed to feature a (close to) constant energy during one microphone rotation. Therefore, if we perform amplitude unmodulation correctly, we expect the corresponding signal to feature constant energy during one microphone rotation, whereas if amplitude modulation remains, we expect fluctuations in energy. This concept is illustrated in Fig. 3, where a 2 kHz sine wave, which has been captured by a rotating microphone, was unmodulated for 360 equally spaced guesses of $\varphi \in [0, 2\pi)$ and the energy of each power spectral density (PSD) frame is plotted for one microphone rotation. As it can be observed, the energy is constant at $\varphi = \pi$, therefore it can be concluded that this is the best DOA guess. Algorithmically, this guess is identified by finding the unmodulated signal with the lowest variance in energy.

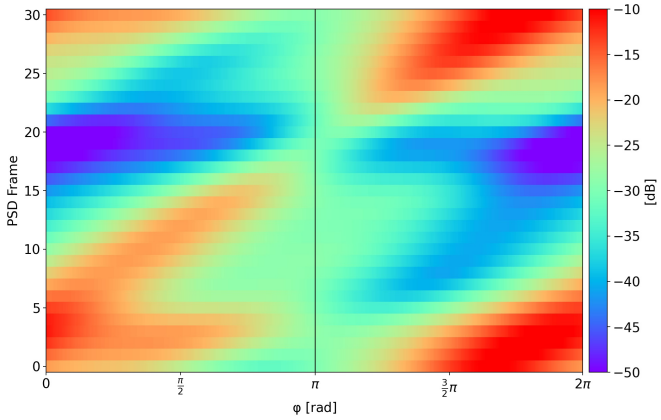



**Figure 3:** Energy fluctuation of a 2 kHz sine wave captured by a rotating microphone with DOA $\varphi = \pi$ that has been unmodulated for 360 equally spaced DOA guesses for one microphone rotation.

## Quantifying frequency modulation

The presence of frequency modulation is identified using a noteworthy property of sinusoidal frequency modulation. This property can be shown by analyzing the PSD of the signal from (2) for different values of the modulation index $\beta$. An example plot of the $\beta$-dependent PSD of (2) for $f_{\text{src}} = 8\,\text{kHz}$, $f_{\text{rot}} = 42\,\text{Hz}$ and $\varphi = \phi_0 = 0$ is depicted in Fig. 4. Note that the PSD was estimated by

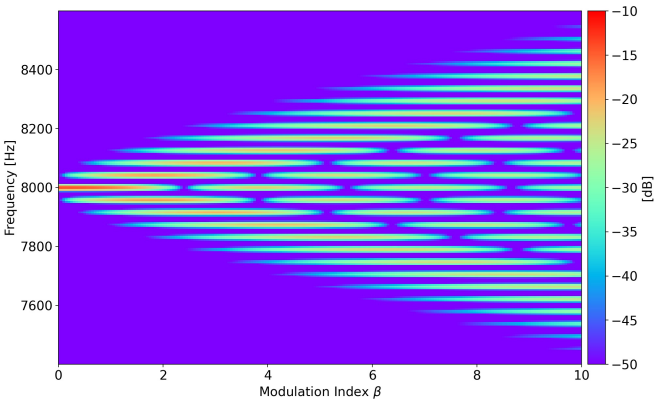



**Figure 4:** PSD of a sinusoidally frequency modulated 8 kHz sine wave for an increasing modulation index.

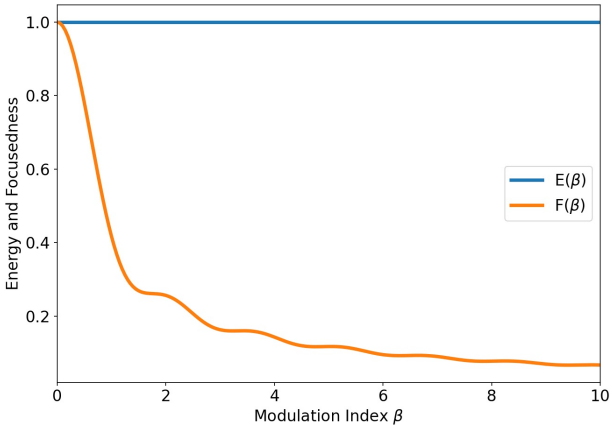



**Figure 5:** Energy and Focusedness of Fig. 4. Both plots have been normalized by dividing each graph by its respective maximum value.

computing the squared magnitude of the DFT of (2), where a DFT length of 8192 was utilized. It can be observed that as the modulation index increases, more sidebands appear around the 8 kHz source signal. This phenomenon is explained by the following alternative formulation of sinusoidal frequency modulation

$$\begin{aligned} &\cos(2\pi f_{\text{src}} t + \beta \sin(2\pi f_{\text{rot}} t - (\varphi - \frac{\pi}{2})) + \phi_0) \\ &= \sum_{n=-\infty}^{\infty} J_n(\beta) \cos(2\pi (f_{src} + n f_{\text{rot}})\, t - n(\varphi - \frac{\pi}{2}) + \phi_0) , \end{aligned} \tag{10}$$

where $J_n(\cdot)$ denotes the Bessel function of the first kind for integer order $n$. Since $\sum_{n=-\infty}^{\infty} J_n^2(x) = 1$ for $x \geq 1$, the energy of (2) remains constant regardless of the modulation index. This is indicated in Fig. 5, where $E(\beta)$ corresponds to the $\beta$-dependent energy of (2). Since the energy is most focused at $\beta = 0$, we introduce a measure called *focusedness*, given by

$$F(\beta) = \sum_{k=0}^{N/2} (S_{xx}(\beta)_k)^2 , \tag{11}$$

where $S_{xx}(\beta)_k$ denotes the $k$-th frequency bin of the PSD of (2). It can be observed from Fig. 5 that this measure decreases as the modulation index increases.

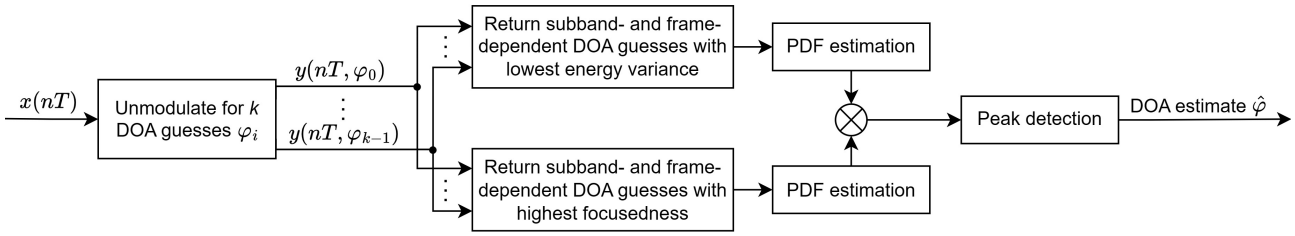
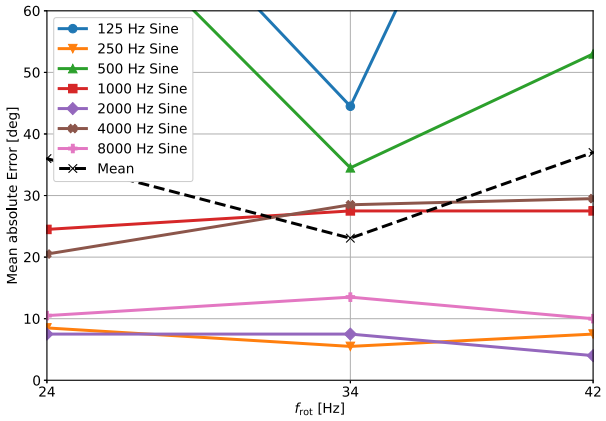
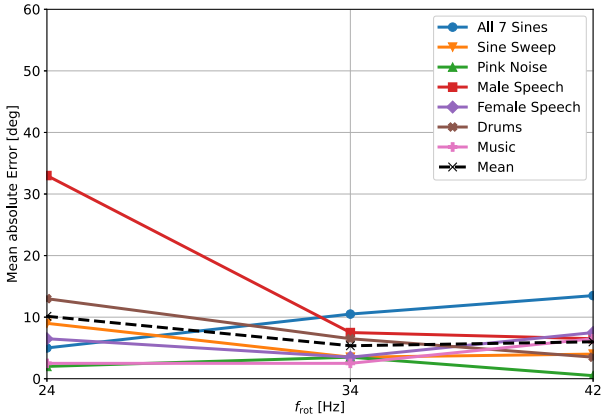



**Figure 6:** Block diagram of the DOA estimation algorithm.




(7a) Results for single-frequency signals.




(7b) Results for wideband signals.

**Figure 7:** Real-world localization accuracy.

## DOA estimation algorithm

These measures now allow for DOA estimation by first unmodulating a given signal for an arbitrary number of DOA guesses and subsequently dividing each unmodulated signal into multiple subbands and frames. Then, for each frame and subband, the signal with the lowest energy variance and highest focusedness is identified and the corresponding DOA is returned. These DOA estimates are subsequently combined by applying probability density function (PDF) estimation and then multiplying these PDFs together. The peak in the final PDF is the final DOA estimate. Fig. 6 depicts the full algorithm.

## Practical experiments and results

The DOA estimation approach was verified using the REM placed at a height of 1.2 m in a low reverberant room with dimensions W×L×H=2.75 m×2.5 m×2.4 m. The utilized test signals correspond to single-frequency sinusoids ranging from 125 Hz to 8 kHz as well as wideband signals such as a sine sweep, music, and speech signals. The results are displayed in Fig. 7. It can be observed that the best performance is achieved at 34 rotations per second and that performance is significantly better for wideband signals. The mean absolute error for single-frequency and wideband signals corresponds to 23 degrees and 5.4 degrees, respectively.

## Conclusion

We presented a novel method for DOA estimation using a single moving microphone. The results indicate that the algorithm performs well in practice, with particularly strong performance for wideband signals. The proposed localization approach may be beneficial in settings involving rotating elements, e.g., integration in Lidar sensors.

## References

[1] Saxena, A., and Ng, A. Y.: Learning sound location from a single microphone. 2009 IEEE International Conference on Robotics and Automation (2009), 1737–1742. doi:10.1109/ROBOT.2009.5152861

[2] El Badawy, D., and Dokmanić, I.: Direction of arrival with one microphone, a few legos, and non-negative matrix factorization. IEEE/ACM Trans. Audio, Speech, Lang. Process. 26 (2011), 2436–2446. doi:10.1109/TASLP.2018.2867081

[3] Fuchs, A. K., Feldbauer, C., and Stark, M.: Monaural sound localization. in Proc. Interspeech (2011), 2521–2524. doi:10.21437/Interspeech.2011-645

[4] Schasse, A., Tendyck, C., and Martin, R.: Source localization based on the Doppler effect. in Proc. of the 13th International Workshop on Acoustic Signal Enhancement (IWAENC) (2012).

[5] Hioka, Y., Drage, R., Boag, T., and Everall, E.: Direction of arrival estimation using a circularly moving microphone. in Proc. of the 16th International Workshop on Acoustic Signal Enhancement (IWAENC) (2018), 91–95. doi:10.1109/IWAENC.2018.8521297

[6] Lawrence, J., Ahrens, J., and Peters, N.: Comparison of position estimation methods for the rotating equatorial microphone. in Proc. of the 17th International Workshop on Acoustic Signal Enhancement (IWAENC) (2022). doi:10.1109/IWAENC53105.2022.9914776

[7] Lawrence, J., Ahrens, J., and Peters, N.: Direction of arrival estimation using the rotating equatorial microphone. Frontiers in Signal Processing (2024). doi:10.3389/frsip.2024.1341087