# Acoustic Localization of Emergency Vehicle Sirens in Traffic

Mattes Ohlenbusch[1,2], Christian Rollwage[1], Joerg Bitzer[1,2]

[1] *Fraunhofer IDMT / Hör-, Sprach- und Audiotechnologie; 26129 Oldenburg, Deutschland*

[2] *Jade Hochschule Oldenburg, Institut für Hörtechnik und Audiologie, 26129 Oldenburg, Deutschland*

*Contact: mattes.ohlenbusch@idmt.fraunhofer.de*

## Introduction

Acoustic source localization methods consist of algorithmic, stochastic or data-driven methods for estimating the position or direction arriving sound originated from. It is utilized in a number of assistive systems, such as hearing aids [1], acoustic monitoring or voice control for ambient assisted living environments [2, 3].

In addition, numerous applications in fields like e.g. robotics [4] or in-car communication [5] are conceivable. In traffic situations where emergency vehicles are deployed, fast reactions and collaboration from traffic participants are critical. For human drivers, the approach of vehicles with high priority is usually signaled visually and by the use of acoustic sirens, of which the latter may be perceived even before a line of sight exists. For vehicles whose driving process is automated to some degree, this may pose a problem since an automated reaction in this situation depends on different factors, for example the direction an emergency vehicle is approaching from.

Although sirens exhibit a characteristic harmonic structure in spectral domain, the ratio between the power of individual harmonics varies with distance, orientation and direction of the vehicle. In addition, the well-known phenomenon of the Doppler-Effect is observed as a frequency shift, depending on the relative position and velocity of both the emergency and the observing vehicle. These effects are visualized in Figure 1.

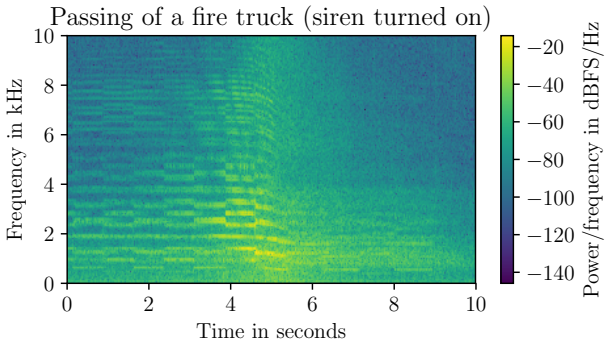


**Figure 1:** Spectrogram of a German firefighter siren in relatively quiet environment. The vehicle is first moving towards, then away from the observer on a straight path. The observer position is fixed throughout the recording.

Traffic situations in which siren localization can be considered critical may display adverse acoustic conditions. These make it critical for potential algorithms to be robust against noise interference which appear in everyday traffic. In this work, acoustic source localization methods for integration in autonomous or semi-autonomous driving are investigated and compared.

## Methods

Various methods were considered to be applied for siren localization. This work draws comparisons between three algorithms, which are described in the following.

## Signal Transfer Model

Usually localization is carried out using arrays of multiple microphones, so that resulting digital signals are discrete with a sample index $k$ and consist of multiple channels. Signals are denoted as $x_m(k)$, where $m$ is an index of the $m$th out of $M$ microphones. Their corresponding Short Time Fourier Transform (STFT) spectra will be denoted as $X_m(n)$ where $n$ is the discrete STFT frequency bin index. Based on STFT spectra, recursive Welch periodograms of power spectral densities are written as matrix $\mathbf{\Phi}(n)$ of dimensions $M \times M$ for each bin.

Under the assumption of additive uncorrelated noise, recorded signals $x_m(k)$ feature a channel-specific noise component and an instance of the useful, correlated signal with a direction- and microphone-dependent delay. As a consequence, many localization algorithms compute correlation-related estimates.

## Steered Response Power

The Steered Response Power (SRP, [6]) is a popular localization method which has already been applied to different environments (e.g. in-car speaker localization [5]). The algorithm is based on delay-and sum-beamforming, in such a way that for different locations or directions, the output of beamformers steered towards them is computed. Steering is done by applying individual time delays to array channel inputs. The SRP algorithm can be formulated as

$$P_{\mathrm{SRP}}(\theta) = \sum_{(m_1,m_2)} \frac{1}{N} \sum_{n=0}^{N-1} \Psi(n) \Phi_{m_1,m_2}(n) e^{\left(j2\pi \frac{n}{N} \tau_{m_1,m_2}(\theta)\right)}. \tag{1}$$

with indices $n$ for one of $N$ STFT-bins and $\theta$ for an individual direction. $\tau_{m_1,m_2}(\theta)$ describes the compensation time delay for microphone $m_2$ relative to microphone $m_1$ associated with the direction $\theta$, and $\Phi_{m_1,m_2}(n)$ the cross-power spectral density of the corresponding signal channels, which in practice has to be estimated. As tuples

$(m_1, m_2)$ only valid unique combinations where $m_1 \neq m_2$ are considered. Frequency-domain weighting can be applied to individual bands as $\Psi(n)$, where $\Psi(n) = 1$ produces a non-weighted estimate. Assuming a single source, a prediction of its position is computed as the maximum value of $P(\theta)$.

## Weighted Phase Transform

A popular method consists of combining the SRP with the Phase Transform (PHAT) [7], which assigns a value in range $[0, 1]$ to each frequency bin. It is defined as

$$\Psi_{\mathrm{PHAT}}(n) = \frac{1}{|\Phi_{m_1,m_2}(n)|}. \quad (2)$$

This is especially adequate for localization of broad-band signals where the spectral power composition varies. Since siren emissions only consist of few frequency components, a modified weighting function $\Psi_{\mathrm{HORN}}(n) = G(n) \cdot \Psi_{\mathrm{PHAT}}(n)$ is considered, where $G(n)$ is a gain component based on a siren denoising algorithm.

## Diagonal Unloading Beamforming

An alternative formulation of the compensation time delay is a steering vector

$$\mathbf{a}(n, \theta) = \left[1, e^{\frac{j2\pi n}{N} \cdot \tau_1(\theta)}, \ldots, e^{\frac{j2\pi n}{N} \cdot \tau_{M-1}(\theta)}\right]^T, \quad (3)$$

with delay $\tau_m(\theta)$ being the delay for microphone $m$ relative to the reference microphone $m = 0$. Using this vector, the Diagonal Unloading beamforming algorithm for source localization [8] is defined. Its output is computed as

$$P_{\mathrm{DU}}(n, \theta) = \frac{1}{\mathbf{a}^H(n, \theta)(\mathrm{tr}(\mathbf{\Phi}(n))\mathbf{I} - \mathbf{\Phi}(n))\mathbf{a}(n, \theta)}, \quad (4)$$

where $\cdot^H$ denotes the conjugate transpose, $I$ an $M \times M$ identity matrix and $\mathrm{tr}(\cdot)$ the trace operator.

## Frequency Fusion

In order to produce an estimate from narrowband beamformer output, a frequency fusion algorithm is necessary. In addition to a simple arithmetic mean (as in equation 1), other methods can be considered, such as the normalized arithmetic mean (NAM) [9]. It is defined as

$$P_{\mathrm{NAM}}(\theta) = \sum_n \frac{P(n, \theta)}{\max_\theta[P(n, \theta)]}. \quad (5)$$

Again, a localization estimate can then be computed as the maximum value of $P_{\mathrm{NAM}}(\theta)$.

## Support Vector Localization

Instead of explicit algorithmic formulations, data-driven approaches to localization have also been proposed. In a Support Vector Machine (SVM) classification approach [1], angular regions are assigned to classes of width equal to the sampling range of beamforming-based approaches. Using training data generated from microphone array impulse responses, a model is fit to classify signal blocks into spatial regions based on pairwise PHAT-weighted cross-correlations. More detailed information on the procedure is found in [1]. In this document, instead of PHAT a modified weighting was used (see section ).

## Evaluation Methods

In order to evaluate siren localization methods, experiments with real traffic noise recordings were carried out. Recordings of stationary siren signals from directions in 45°-steps in the azimuthal plane were added. As evaluation measures, the root mean squared error (RMSE) and the accuracy were computed. In this context, accuracy is defined as the proportion of estimates that were within 15° range of the correct direction.

In addition to that, an experiment with stationary sirens in increasing source-receiver-distance were executed to investigate the robustness of localization methods for distant sirens.

A third experiment featured recordings from a fire truck passing by a microphone array, with the siren turned on continuously. Since using alarm sirens without permission, e.g. in an emergency situation, is against the law in Germany, these recordings were created on a former air field with the assistance of local firefighters. Figure 2 illustrates the recording procedure. All recordings were

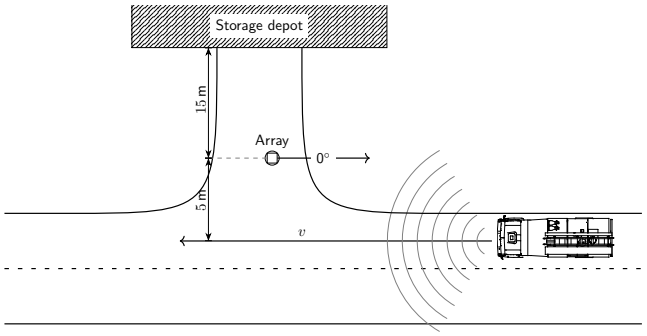


**Figure 2:** Schematic illustration of the experiment carried out in order to evaluate algorithmic performance for a moving siren source.

created using an eight-microphone array also described in [5].

## Results and Discussion

Localization performance measures resulting from the experiments with non-moving sirens and additive environmental noise from a traffic-reduced zone and a busy town square are shown in Tables 1 and 2, respectively. In both situations, considered approaches show a very high accuracy, with the RMSE values being slightly lower in the town square situation. It should be noted that while the spectral composition of both the target signal and the environmental noise show characteristics featured in real applications as well, the added difficulty of moving

**Table 1:** Experimental results on localization of non-moving sources in environmental noise recorded in a traffic-calmed zone.

| Method | DU | SRP | SVM |
|---|---|---|---|
| Pre-/Postprocessing | NAM | HORN | HORN |
| RMSE (degrees) | 0.000 | 0.184 | 3.626 |
| Accuracy | 1.000 | 1.000 | 1.000 |

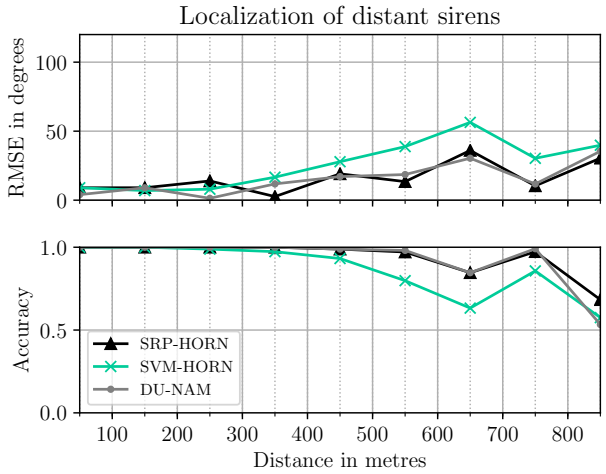


**Figure 3:** Localization results for a standing emergency vehicle siren over distance to receiver. Measurements were taken approximately every 100 m, starting at 50 m.

sources and receiver vehicle are neglected. Another simplification is that siren recordings used in this experiment were taken from close proximity, while in reality emergency vehicles should be locatable from distance already.

**Table 2:** Experimental results on localization of non-moving sources in environmental noise recorded at an active town square.

| Method | DU | SRP | SVM |
|---|---|---|---|
| Pre-/Postprocessing | NAM | HORN | HORN |
| RMSE (degrees) | 9.729 | 0.737 | 21.107 |
| Accuracy | 0.984 | 1.000 | 0.973 |

Figure 3 shows localization RMSE and accuracy for the second experiment investigating robustness against distance. As is to be expected, localization error overall increases with distance. As also observed in Figure 1, the power of high frequency siren overtones decreases more than for low frequencies, which in turn may diminish the effectiveness of broadband weighting methods. In this experiment, DU-NAM and SRP-HORN methods perform equally well, while using the SVM-HORN algorithm results in lower accuracy and error.

The third experiment was carried out without measuring ground truth direction of arrival-values. Following the description of the recording process in section  however, it can be followed that the vehicle starts approaching the array from the front, defined here as 0°. The rate of directional change should increase the closer the fire truck is to the array, and after passing by this trend should reverse, with the vehicle ending up behind the array at approximately 180°. Since no ground truth data is available, only qualitative observations can be made. In Figure 4, results for individual STFT blocks computed with the considered methods are displayed. Overall, all three subplots feature the characteristic progression expected from the recording procedure. It should be noted, however, that the SVM-HORN estimates feature a spread in the beginning of the recording, where the vehicle is far away from the receiver. This observation is consistent with results from the previous experiment. SRP-HORN and DU-NAM results also feature outliers, but in smaller numbers. During the event of the truck passing the array position at around 30 seconds, all three methods result in the temporal progress expected from the experimental design.

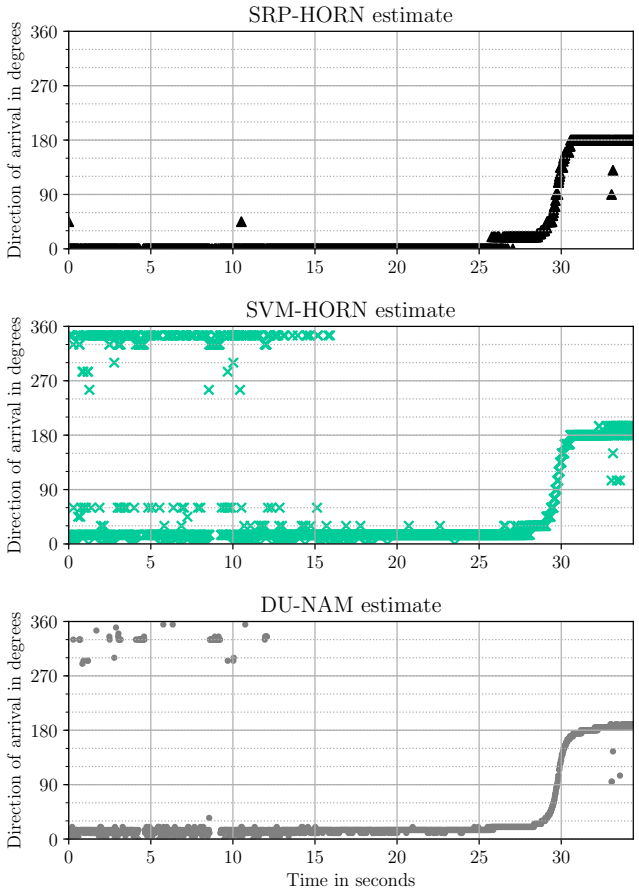


**Figure 4:** Block-wise localization results using a recording of a fire truck passing by at a constant speed of $v = 100\,\mathrm{km/h}$, in the scenario outlined in Figure 2.

## Conclusions and Outlook

The experiments previously described and evaluated indicate that localization of sirens in traffic situations is viable. Of the algorithms utilized, the SRP-HORN and the DU-NAM algorithm are considered to be particularly suited for this task, as they were found to be more robust against both traffic noise and distance effects on siren signals. From the experiment featuring a moving source, promising results were gathered as well. Still, it remains to be investigated how effects from a moving receiver position and corresponding engine sounds from a fixed direction relative to the array influence localization, and how these influences may be mitigated. In order for the problem of siren localization to be solved in a practical way, object tracking methods have to be considered, as temporal information can be be used to increase robustness against outliers as observed in the third experiment. Considering the fact that oftentimes larger emergency operations feature multiple vehicles, practical implementations of siren localization may need to employ multi-object tracking solutions in order to provide reliable results.

## Acknowledgments

The authors would like to express their gratitude to the members of the *Freiwillige Feuerwehr Metjendorf* for their major support during the realization of the experiments.

## References

[1] Hendrik Kayser and Jörn Anemüller. "A discriminative learning approach to probabilistic acoustic source localization". In: *2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE. 2014, pp. 99–103.

[2] Christian Bartsch et al. "Evaluation of different microphone arrays and localization algorithms in the context of ambient assisted living". In: *International workshop on acoustic echo and noise control*. 2010.

[3] Stefan Goetze et al. "Acoustic monitoring and localization for social care". In: *Journal of Computing Science and Engineering* 6.1 (2012), pp. 40–50.

[4] Laurent Calmes, Gerhard Lakemeyer, and Hermann Wagner. "Azimuthal sound localization using coincidence of timing across frequency on a robotic platform". In: *The Journal of the Acoustical Society of America* 121.4 (2007), pp. 2034–2048.

[5] Mattes Ohlenbusch et al. "Evaluation of Speaker Localization methods for Vehicle Interior Applications". In: *45th Deutsche Jahrestagung für Akustik (DAGA'19), Rostock, Germany* (2019).

[6] Joseph H DiBiase, Harvey F Silverman, and Michael S Brandstein. "Robust localization in reverberant rooms". In: *Microphone Arrays*. Springer, 2001, pp. 157–180.

[7] Charles Knapp and Glifford Carter. "The generalized correlation method for estimation of time delay". In: *IEEE transactions on acoustics, speech, and signal processing* 24.4 (1976), pp. 320–327.

[8] Daniele Salvati, Carlo Drioli, and Gian Luca Foresti. "A low-complexity robust beamforming using diagonal unloading for acoustic source localization". In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 26.3 (2018), pp. 609–622.

[9] Daniele Salvati, Carlo Drioli, and Gian Luca Foresti. "Incoherent frequency fusion for broadband steered response power algorithms in noisy environments". In: *IEEE Signal Processing Letters* 21.5 (2014), pp. 581–585.