

# Simulation framework for detecting and tracking moving sound sources using acoustical beamforming methods

Péter Tapolczai, Péter Fiala, Gergely Firtha, Péter Rucz

*Laboratory of Acoustics and Studio Technologies,*

*Budapest University of Technology and Economics, H1117 Budapest, Hungary, Email: tapolczai.peter@simonyi.bme.hu*

## Introduction

Microphone arrays and beamforming procedures enable the localization of sound sources and are used in a great number of industrial applications. Remote sound sources can be localized by creating an acoustical image of the spatial distribution of the source strength. Acoustical focusing can also be implemented in order to attain the filtered signal of selected sources. Application to moving sound sources involve a number of challenges, such as taking the Doppler shift into account, or to apply the acoustical focusing with a time-varying focal point.

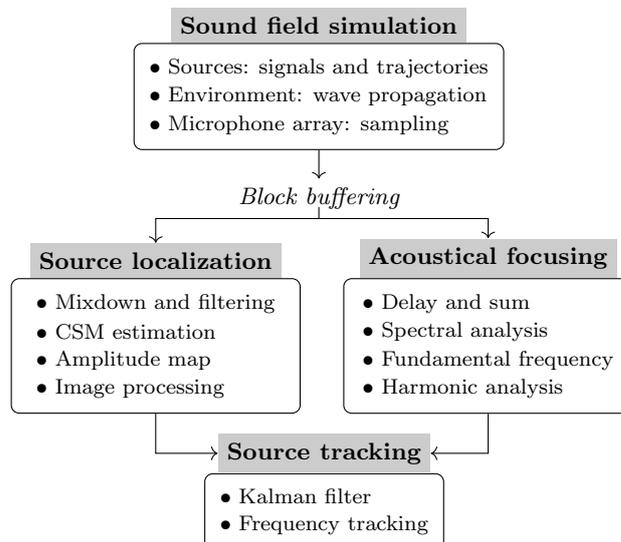
In this contribution a simulation framework for the detection and tracking of moving sources is introduced. It is shown that in case of moving sources, source localization and acoustical focusing can be implemented in a feedback loop to enhance the quality of the detection. The object oriented simulator is capable of evaluating the sound field of arbitrary moving sound sources and contains implementations of various beamforming methods. The reconstruction of the trajectory of the moving source is supplemented by a nonlinear Kalman filter. The tracking of a small unmanned aerial vehicle is demonstrated as an example application.

## Simulation framework

The framework introduced in this paper is a MATLAB-based object-oriented toolbox. The simulator serves two main purposes. On the one hand, it allows for the simulation of scenarios, involving multiple moving sound sources and arbitrary microphone array configurations. On the other hand, it provides a unified structure for testing beamforming and source localization algorithms.

The sound field simulation module is capable of reproducing the sound signals at stationary receiver locations (i.e., microphone positions), created by a custom number of sound sources moving along arbitrary trajectories. In order to properly account for the movement of the sources the emission times are calculated at each time instance for all source–receiver pairs.

The signal processing tools enable the acoustical focusing to the moving sources using the delay and sum method. The framework currently contains implementations of the CLEAN-PSF, CLEAN-SC [1], DAMAS [2], and MUSIC [3] algorithms beside the conventional beamforming method for acoustical imaging and subsequent source localization. By means of a Kalman filter, source localization estimations can be averaged and fused with



**Figure 1:** Structure and data flow of the framework.

predictions from a dynamical model. Finally, in case of sound signals with significant harmonic content, such as the noise radiated by rotating propeller blades of small unmanned aerial vehicles (UAV), a frequency tracking algorithm can supplement localization and source tracking.

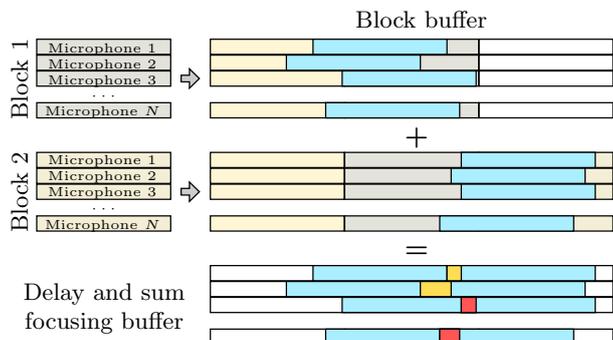
Figure 1 displays the structure and the data flow of the simulation framework. The next sections introduce and demonstrate the signal processing parts of the simulator, i.e., acoustical focusing, source localization, and tracking.

## Focusing to moving sound sources

In case of moving sources, the acoustical camera is focused to the target in each block with different phase shifts. As a result, consecutive blocks can overlap or a gap can occur between the segments, as illustrated in Figure 2.

Two methods are implemented in the framework to solve this problem. One method is to interpolate each block to the length of the path in the previous cycle by stretching its time scale, so that the gaps disappear. The advantage of this solution is that it also takes the speed of the source into account and changes the size of the current block, so it can compensate the change in frequency shifts resulting from the Doppler effect to some extent. However, it has the disadvantage that it is computationally intensive, so its real-time utilization is limited.

Alternatively, an additional block is calculated between each pair of blocks, focused to an interpolated position



**Figure 2:** Block summation problem. Different delays in two consecutive blocks (light blue bricks) can result in gaps (yellow block parts) or overlapping segments (red block parts).

between the two samples, each shifted block being weighted using a window function, and the block summation is then performed. The advantage of this method is that it is much less involved computationally, compared to the previous approach. Similar windowing methods are used in lossy audio compression, where the windowing is required to smooth the noise resulting from quantization with different number of bits in adjacent frames.

During testing, no significant differences were observed between the focused signals resulting from the two methods, the received signal was free of gaps and overlapping artefacts in both cases.

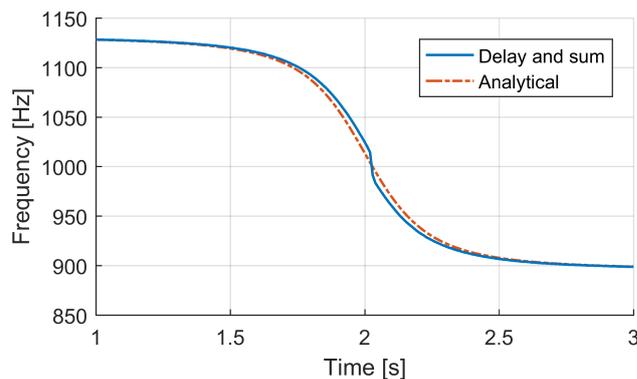
### Examination of the Doppler effect

In the case of moving sources, the Doppler effect can have a significant influence on the frequency of the received signal. In our case, this is also interesting because the amplitude map is evaluated in the vicinity of a predefined nominal frequency.

In this simulation case, the Doppler frequency shift of a signal resulting from acoustical focusing is examined. The sound field is sampled by using a circular microphone array, consisting of 24 microphones and having a radius of 1 m. The source travels along a straight line at a constant speed of 40 m/s, parallel to the plane of the microphone array at 10 m distance and emits a harmonic signal with 1 kHz frequency. Acoustical focusing is performed by applying a block length of 0.01 s, with the moving focal point set as the position of the source at the emission time corresponding to the middle of the block.

The analytically calculated frequency and the measured frequency of the focused signal are compared in Figure 3. The results nearly coincide and the frequency of the signal decreases continuously, as expected.

Note that in Figure 3 the frequency of the focused signal is slightly above the analytical frequency until the source passes the microphone array. Then, a small jump is observed, and when the source moves away from the microphone array, the measured frequency is lower than the analytical one. The phenomenon is explained by the fact that while in the analytical case the frequency was calculated by using a single receiver at the center of the



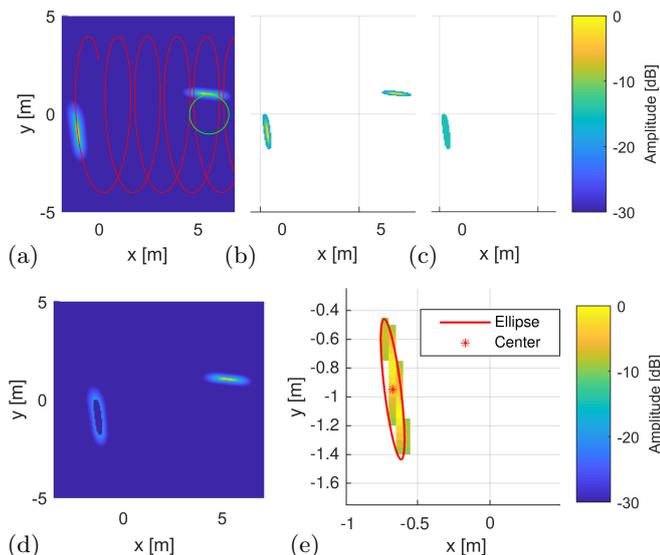
**Figure 3:** Effect of acoustical focusing on the measured frequency of a moving harmonic source.

microphone array, the actual microphone array has a finite extent. In the vicinity of the jump, the source is still approaching some of the microphones, while it is already receding from the others. At the point of the frequency jump, the total weights of these two groups of microphones change rapidly. It was observed that a larger array results in a greater frequency jump.

### Source localization

Using the microphone array data, an acoustical image is created by the conventional beamforming method or an image-cleaning algorithm (MUSIC, DAMAS, CLEAN). As these methods operate in the frequency domain, the signals are band-filtered and the cross spectral matrix (CSM) is estimated. The imaging algorithms use the CSM for computing the amplitude map. Once the acoustical image is created, it is necessary to estimate the location of the sources in the observation area. It is also desirable to quantify the uncertainty (e.g. the covariance matrix) of the estimated locations.

To estimate the source locations, the acoustical image is processed in a cyclical manner. Our localization algorithm repeats the sequence of internal steps until the amplitude is above a certain threshold. The steps of one cycle are illustrated in parts (a)–(e) of Figure 4. At the beginning of each iteration, the amplitude map (a) is cleared where its values are under a given threshold level (b). Then, a so-called connectivity analysis is carried out on this modified amplitude map, and hence the largest contiguous object is found (c). It is necessary to “delete” the given source from the amplitude map, which is done by resetting the amplitude map to a lower threshold in a wider contiguous area around the found source (d), such that in the next cycle the same source is definitely not found again. Finally, an enclosing ellipsoid is matched to the contiguous object (e) in order to estimate the statistics of localization: the center of the ellipsoid indicates the expected value of the distribution, and the lengths and angles of the axes are related to the covariance matrix. The steps of fitting the minimum volume enclosing ellipsoid (MVEE) are described in [4]. Then, in the next cycle, the same steps are performed anew, until there is no new source found.



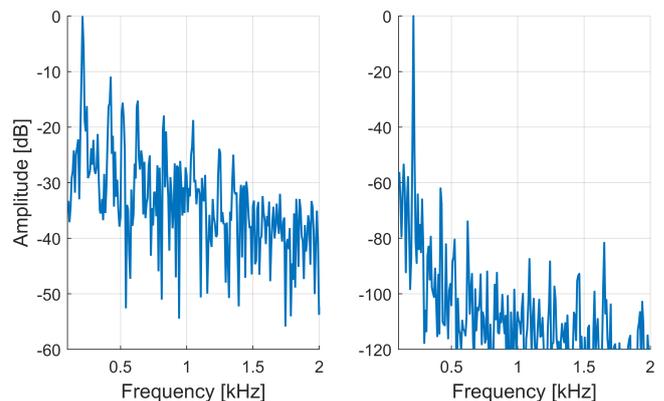
**Figure 4:** Illustration of the source localization steps in case of two moving objects. (a) Detection amplitude map. (b)–(e) Localization steps for the first source.

### Tracking supported by Kalman filter

The Kalman filter provides an optimal state estimate of an object by recursively averaging the noisy input data and combining the measurements with estimations based on the dynamic model of the object. One useful feature of the Kalman filter is that it also estimates the covariance matrix of the estimated state vector.

The filter is optimal, if the dynamics of the system are linear and the noise is additive with a Gaussian distribution. However, if any of these conditions does not hold, the conventional Kalman filter can no longer be used, and extensions are required. Extended Kalman Filters (EKF) are based on some kind of linearization, requiring the evaluation of the Jacobian matrix, which can be computationally involving. As an alternative, scattered state variables can be generated around the estimated state vector and the statistics are evaluated by transforming these scattered states using the non-linear system or output equation, respectively. A special selection of scattered states is used by the Unscented Kalman Filter (UKF) that calculates the transformed statistics using so-called sigma points that are located on the standard deviational ellipsoid. Hence, the number of scatter points is well defined and the computational costs are moderate. Moreover, it can also be shown that even in the case of multiple nonlinear transformations, UKF’s sigma point choice is optimal [5].

In our framework, a nonlinear relationship exists between the observation variables, being spherical coordinates  $(r, \vartheta, \varphi)$ , and the state variables that contain the position and velocity of the tracked object in the Cartesian coordinate system. This choice is explained as follows: in case of infinite focal length, no information is given about the distance  $r$ , while for finite focal lengths, it can reasonably be assumed that the distance estimation has different accuracy than that of the angles.



**Figure 5:** Left: Amplitude spectrum of a quadcopter recorded in a semi-anechoic chamber at  $\approx 6000$  RPM. Right: Result of HPS calculation.

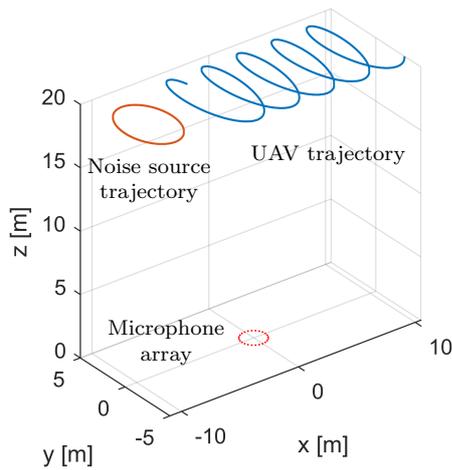
The dynamic model of the source is described by assuming uniform motion, i.e.,  $\dot{\mathbf{x}} = \mathbf{v}$  and  $\dot{\mathbf{v}} \approx \mathbf{0}$ , with  $\mathbf{x}$  and  $\mathbf{v}$  denoting the position and the velocity of the object. This system of equations is discretized in time, with the sampling interval  $T_s$  being the time difference between two consecutive source localization steps. Naturally, the source is not expected to follow a linear trajectory. This is taken into account by assuming a large system noise for the  $\dot{\mathbf{v}} \approx \mathbf{0}$  equation, resulting in a velocity estimation relying far more on the measured positions than the model prediction. If a position estimate is not available, i.e., the source was not found on the acoustical image, uniform motion is predicted with a steeply increasing variance of the estimated state.

### Example case: tracking a simulated UAV

In order to supply more realistic input signals for the simulation framework, the sound of a DJI-F450 quadcopter was recorded in a semi-anechoic chamber under static (non-flight) conditions with different propeller configurations and at various rotor speeds. The directivity of the sound radiation was also measured. It was found that assuming a uniform directivity is a satisfactory approximation in case of the small quadcopter, thus, this assumption was used in the simulation environment.

The left diagram of Figure 5 shows the amplitude spectrum of the recorded sound with four active rotors at  $\approx 6000$  RPM. The blade passing frequency (BPF) is  $\approx 200$  Hz in this case. As observed, there is significant harmonic content in the received signal, emerging from the broadband noise. The harmonic content can also be exploited in the detection and localization of the source. In order to detect the BPF, a frequency detection routine, such as the Harmonic Product Spectrum (HPS), can be applied. The right diagram of Figure 5 shows the result of the HPS calculation and demonstrates the achievable gain with respect to the signal to noise ratio.

Figure 6 depicts the simulation arrangement in which a moving UAV is tracked. A moving broadband noise source disturbs the detection and localization of the object. The simulated quadcopter emits the recorded



**Figure 6:** Arrangement for tracking a simulated quadcopter. A broadband noise source disturbs the detection.

sound of the real copter and moves with a constant velocity magnitude of  $|\mathbf{v}| = 20$  m/s on a spiral trajectory. The sound field is sampled by a circular microphone array with the radius of 1 m and consisting of 24 sensors.

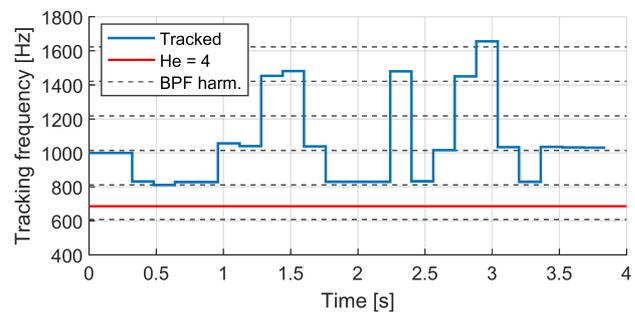
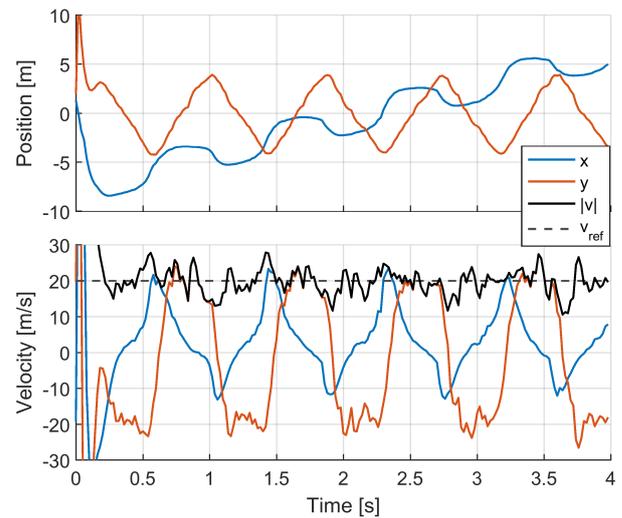
The results of the source tracking are displayed in Figure 7. As it is depicted, the Kalman filter gives a reliable estimate of both the position and the velocity from mere position measurements. During the initial settling time of the filter, the predicted covariance decreases steeply. After settling, the estimated variances of the state variables are in good correspondence with the real statistics.

The bottom diagram of Figure 7 shows the result of the frequency tracking algorithm. When the BPF is determined using the HPS method, a harmonic to be tracked is selected. One of the criteria for this selection is due to the applied MUSIC algorithm, which is only effective in case of Helmholtz numbers  $He > 4$  [6]. The other requirement is that the signal to noise ratio should be high in the neighborhood of the selected harmonic. As seen, the frequency tracking algorithm adaptively selects a harmonic of the estimated BPF, depending on the temporal changes of the SNR. The selected frequencies are close to the harmonics of the theoretical BPF. By using the frequency tracking based on the acoustical focusing to determine the source localization frequency, and using the localization output to set the focal point of focusing, a feedback loop is introduced in the tracking mechanism.

This example demonstrates the capability of the simulation framework for following a moving sound source in a noisy environment. The focusing, localization, and source tracking algorithms will be tested on data from real measurements in the near future. Finally, techniques for acoustical source classification will be investigated.

## Acknowledgments

This project is supported by the European Union, project ref.: GINOP-2.2.1-15-2017-00087. P. Rucz acknowledges the support of the Bolyai János research grant provided by the Hungarian National Academy of Sciences.



**Figure 7:** Tracking of a moving quadcopter. Top: position and velocity from the UKF. Bottom: frequency tracking.



Supported by the ÚNKP-18-4 New National Excellence Program of the Ministry of Human Capacities.

## References

- [1] P. Sijtsma. CLEAN based on spatial coherence. Technical Report NLR-TP-2007-345, National Aerospace Laboratory NLR, 2007.
- [2] T. F. Brooks and W. M. Humphreys. A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays. *Journal of Sound and Vibration*, 294:856–879, 2006.
- [3] H. L. van Trees. *Optimum array processing*, pages 1158–1163. Wiley & Sons, New York, 2002.
- [4] S. Bonnet, C. Bassompierre, C. Godin, S. Lesscq, and A. Barraud. Calibration methods for inertial and magnetic sensors. *Sensors and Actuators A:Physical*, 156(2):302–311, 2009.
- [5] D. Simon. *Optimal state estimation – Kalman,  $H_\infty$ , and nonlinear approaches*. John Wiley and Sons, New York, 2006.
- [6] G. Herold and E. Sarradj. Performance analysis of microphone array methods. *Journal of Sound and Vibration*, 401:152–168, 2017.