STRENGTHS AND WEAKNESSES OF CALCULATING BEAMFORMING IN THE TIME DOMAIN

Olaf Jaeckel

Society for the Promotion of Applied Computer Sciences (GFaI e.V.)
Rudower Chaussee 30, 12489 Berlin-Adlershof, Germany
jaeckel@gfai.de

ABSTRACT

Most of the present beamforming systems in commercial and academic applications are performing their calculations solely in the frequency domain. While the early historical beamformers actually have been implemented in the time domain by inclusion of simple analog hardware delay lines, this principle was not applied anymore due to the availability of digital signal processors which allowed the calculation of fast Fourier transforms already within the data acquisition hardware frontends. Nowadays, modern computers offer the feasibility to process huge amounts of multichannel data in the frequency domain as well as in the time domain, so computational effort is not the only concern anymore.

The paper compares the benefits and the disadvantages of both domains, focusing on the time domain because the latter is neglected or at least heavily underrepresented throughout the classical array signal processing textbooks and most of the current papers dealing with the beamforming method. Therefore, it is not the purpose of this paper to present a new theory or any advanced algorithms, but rather to give a general overview which is motivated from an application-oriented and practical standpoint of view. It is pointed out that a high signal bandwidth and a sampling rate much higher than twice the Nyquist frequency necessary for the channel data are especially important when working in the time domain. Application examples from technical and nontechnical fields will be given.

1   INTRODUCTION

The use of microphone arrays and multichannel data recorders in connection with software for a fast visualization of the results (Acoustic Camera) has become quite popular for the localization of sound sources of machinery and equipment of any kind. In modern state of the art systems, an automatic overlay of an optical photo or an extracted edge detected image of the object with the calculated acoustic colour map is provided which gives the user a fast overview of the dominant noise sources emitted by the device under test. A newly developed system even allows the overlay of a sound pressure map onto the surface of a three dimensional model of the test object.

While the commercially available products and the systems used in academic research may differ in array geometry and channel number, in algorithmic complexity and implementation details as well as in software speed and user comfort, the underlying common principle of all those systems in the farfield approach is the well known "delay-and-sum"-beamforming method. Even though in the meantime some of the nearfield systems (based on nearfield acoustic holography) are also termed "acoustic cameras", we will not consider such nearfield methods in this paper.

The behaviour of the very basic delay-and-sum method when computed in the time domain and in the frequency domain will be compared. Undoubtedly, both domains are completely equivalent in mathematical theory, because the real sound pressure values from the microphones and their complex Fourier transforms represent the same information content. However, from a practical viewpoint and due to different algorithmic properties, there are some quite important differences between both domains which will be explained in the sequel.

2   DELAY-AND-SUM BEAMFORMING – OVERVIEW

2.1 Basic principle

The term "beamforming" historically stems from active localization systems (SONAR, phased RADAR) and denotes the fact that the main lobe (the so called "beam") of the directivity/sensitivity pattern of a discrete sensor array changes its form dependent on the actual method used to virtually steer the array to certain angular directions. But nowadays, the term is used synonymously for active systems as well as for purely passive localization methods.

With the most simple method, classical delay-and-sum beamforming [1], a microphone array will be successively focused to many points lying on a measurement plane or on an object's surface. In theory, this focus distance can even be considered to be infinitely long which is equivalent to the model assumption of plane waves passing through the sensor array. For every individual focus point or direction, the relative runtime delays between the microphone channels are compensated for. The time signals thus shifted will then be added up coherently. Dividing by the channel number gives an estimated time function which is comparable in its power content to the original time signal at this focus location or, in case of infinite focus range, from a certain angular direction.

For each of those many time functions, interesting parameters as sound pressure level can be determined easily and simultaneously. This way, a mapping of the complete sound pressure distribution in the measurement plane or on the surface of an object can be calculated.

2.2 Calculation in the time domain

The simplest approach is the straightforward calculation of a delay-and-sum-beamformer in the time domain. The reconstructed time function at every location **x** is calculated as:

$$\hat{f}(\mathbf{x},t) = \frac{1}{M}\sum_{i=1}^{M} w_i f_i(t - \Delta_i). \tag{1}$$

In Eq. (1), t denotes time, M is the number of microphones in the sensor array, and the w_i are optional spatial shading weights acting similar to the windowing coefficients applied before performing time signal spectral transforms to reduce leakage and smearing effects. But for our Acoustic Camera's standard ring array, all the w_i are simply set to unity because spatial shading does not make sense in this case. Eq. (1) is also illustrated in Figure 1 below.

The $f_i(t)$ are the recorded time functions of the individual microphones, and the Δ_i are the appropriate relative time delays, which are calculated from the absolute run times $\tau_i = |\mathbf{r}_i| / c$ by subtracting the minimum over all τ. The symbol c denotes the speed of sound in air and $|\mathbf{r}_i| = |\mathbf{x}_i - \mathbf{x}|$ is the geometrical distance between the spatial position of microphone number i and the actually calculated focus point **x**.
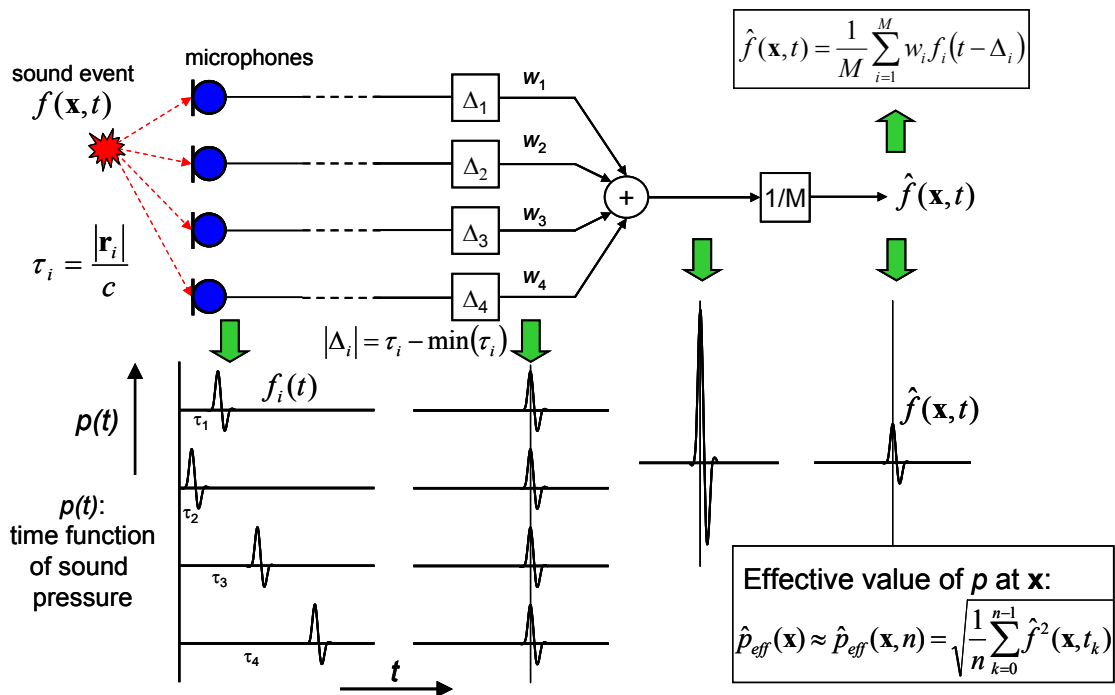

Fig. 1. Basic principle of the delay-and-sum beamformer as calculated in the time domain.

In practice, we prefer this distance calculation because it is equivalent to the use of a spherical wave model and it has proven to be very useful as it automatically and smoothly translates to the above mentioned model of a plane wave for larger focal distances |**r**|. The classic array literature, however, often preferres just the plane wave model.

Despite its extreme simplicity, the delay-and-sum method in the time domain is quite robust and powerful and has shown its practical usability in an extraordinary wide range of acoustic localization and trouble shooting applications for years now.

2.3   Calculation in the frequency domain

Using linearity and the shifting property of the Fourier transform, Eq. (1) can easily be written in the frequency domain. Taking the Fourier transforms of the individual microphone signals now yields the reconstructed spectral function at **x**:

$$\hat{F}(\mathbf{x},\omega) = \frac{1}{M}\sum_{i=1}^{M} w_i F_i(\omega) \cdot e^{-j\theta_i}. \tag{2}$$

In Eq. (2), the $\theta_i = \omega\Delta_i(\mathbf{x})$ denote the location- and frequency dependent phase angles by which the component of every partial frequency ω of the i-th signal has to be shifted in order to exactly compensate for the relative runtime delays $\Delta_i(\mathbf{x})$ of the individual microphone channels in the superposition of all the spectral components.

The terms $exp(-j\theta_i)$ in Eq. (2) dependent on focus direction are usually summarized into one vector that is therefore called the steering vector. For every individual frequency bin, there naturally exists one such steering vector. Therefore it is clear that the evaluation of Eq. (2) for just a single signal frequency (narrowband case) can be computed very fast and easily.

If only the effective signal power is of interest, as is the case for the determination of the sound pressure level at every focus point, this effective value is usually calculated directly in the frequency domain via computation of the complex cross spectral matrix. An inverse transform back to the time domain is not necessary to merely compute acoustic photos.

A first advantage of the frequency domain beamforming is achieved by the transformation of time shifts into phase rotations. The connection of the individual microphone signals with the focus point or directional information now reduces to simple complex multiplications. This decoupling of the microphone signals from the localization information in the frequency domain has another distinct advantage: Here, the Δ_i can be processed as continuous numerical values as actually calculated from the real focus distances which completely avoids the often very disturbing sampling frequency dependent quantization of the steering delays present in the time domain. So, frequency domain beamformers can usually operate at lower sampling rates than pure time domain implementations.

The isolation of the microphone data from the steering vector information also allows the independent use of the complex cross spectral matrix of the microphone signals in advanced signal processing applications. The array signal processing literature describes a wide range of specialized beamforming variants (e.g. adaptive data dependent beamformers or subspace methods based on eigendecompositions) and more sophisticated algorithms in the frequency domain [1], [2], [3]. Many of those advanced methods have originally been developed for the

narrowband case only, and though most of them can be extended to the broadband case by applying a narrowband method repeatedly for every frequency bin or at least for averaged frequency bands, this general case clearly requires substantially more computational effort. The reason for this disadvantage of the frequency domain beamforming is based on the simple fact that even for a single constant time domain shift Δ_i, the necessary phase rotation θ_i will be a different one for every individual frequency component, thus enforcing the separate calculation of another phase angle for every individual spectral line or band. This makes the calculation of Eq. (2) at every focus point to appear as a quite elaborate and tedious work in the practically very important case of broadband acoustic signals.

3   STRENGTHS AND LIMITS OF TIME DOMAIN BEAMFORMING

3.1   Advantages of the time domain approach

It should be noted that the above mentioned restriction to the narrowband case in the spectral domain was just a matter of historical development and not of any theoretical requirements. In the meantime, computers are powerful enough to process broadband systems in the frequency domain as well. On the other hand, this earlier restriction may have caused an unnecessary blocking of the traditional acoustician's view onto the potential advantages of a direct implementation of Eq. (1) in the time domain.

One of the most obvious but seemingly often overseen properties of the time domain beamforming in Eq. (1) is that it inherently constitutes a ready-to-use broadband method. Hence, the calculation in the time domain basically offers great virtues for signals that are coming from broadband sources anyway, as are the most sounds of technical machinery. Those sound sources are only very seldom composed of just a few tonal components, but they are rather to be considered as a mixture of broadband noise and many different tonal signals emitted by rotating components or vibrating structures of the machine under test.

The same advantage of the time domain beamformer primarily holds for short pulses and transient signals of any kind which basically should not be difficult to localize in space but which also exhibit very broad spectra. This makes spectral domain calculations expensive again. Here, the simple time domain formula Eq. (1) easily works without all the pain and overhead that is present in the frequency domain.

Especially for transients, the calculation approach in the frequency domain completely unneccessarily transforms a quite simple problem into a rather complicated one. In short time spectral calculations, first of all time frames and overlapping intervals have to be determined which reduces the achievable time resolution. Using blocks short enough for the transients loses frequency resolution on the other hand. Next, the blocks must be windowed which reduces their actual signal energy. After the Fourier transforms, cross spectral matrices have to be estimated but those estimations are known to be inconsistent especially for short sample lengths. This enforces averaging over many time blocks which again reduces time resolution further and often makes stationarity assumptions, but exactly those assumptions are not fulfilled in this particular case of transient signals. After this estimation, finally the steering information can be multiplied in.

And, not to forget, for a broadband signal this awkward procedure has to be done for every individual frequency bin or band. So, using the frequency domain beamforming is really an unlucky choice for this transient signal type.

Conversely, localizing transient noise sources and instationary signals is a very quick and easy task in the time domain. The full possible time resolution according to the channel sampling rate can be exploited, no further assumptions are made, and no signal energy is lost or averaged out.

Another practically important advantage of the time domain approach is its high computational efficiency. While computing time scales only linearly with channel number in the time domain, this scaling is quadratic with the microphone number in the frequency domain. Calculation time is a very important issue particularly for systems intended for commercial use or for processing huge amounts of data in greater measuring campaigns.

A further positive side effect of the explicit broadband computations is that there are less aliasing problems. Though this effect is more related to the actual signal, array geometry and achievable contrast than to the domain the calculations are performed in, its benefits are more likely to show themselves in the time domain. While a real source at a defined position will have its mainlobe width changed with frequency but not its location in the acoustic photo, the sidelobes do change their locations dependent on signal frequency. With a broadband signal, as a practical result the sidelobes of the individual frequencies would appear at different positions and effectively average out. For a strongly tonal signal, the sidelobes belonging to this particular frequency will be clearly visible in both domains. But strong sidelobes will also be visible for an intrinsic broadband signal that is analyzed with a mere narrowband method in the spectral domain.

A last but not unimportant strength of the timedomain approach is that it is better suited to implement Doppler corrections for fast moving sources, as e.g. reported in [4]. This can be achieved via an adapted focus distance calculation where the focus point in the image field is dynamically following the moving source at known speed, which essentially corresponds to a digital resampling procedure.

3.2   Limitations of the time domain beamformer

Implementation of a beamformer directly in the time domain also has some drawbacks that should be pointed out. One of the main disadvantages is that high sampling rates are inevitable in order to minimize the malicious effects of the discrete steering delay quantizations on sidelobe height and image contrast. A time domain beamformer sampling just at the absolutely necessary Nyquist rate would exhibit bad angular and spatial resolution and strongly changing sidelobe patterns (resulting in a bad signal-to-noise ratio) dependent on array geometry, on focus point and signal frequency. Those effects are known qualitatively, but they can hardly be described analytically safe for very trivial array geometries.

According to our own practical measurement experiences, it is recommendable to sample at least at a rate about ten times higher than the highest interesting signal frequency, so the data recorder of the Acoustic Camera samples at 192 kHz per channel to savely cover the normal audio range. The necessity to sample at high rates needs more memory resources and increases data transmission times, so the hardware requirements are higher.

As a consequence the time domain beamforming also has the need for highly optimized calculation algorithms with respect to memory usage and run time behaviour.

The calculation of spectrally resolved acoustic photos for stationary signals can be performed faster in the spectral domain that only needs the Fourier transforms of the M microphone signals, whereas in the time domain the FFT must be processed afterwards for every reconstructed time function, i.e. for every individual pixel in the acoustic image. This operation scales with the square of the image's edge length. For longer marked time intervals, the memory demand for this computations can become prohibitive in both domains.

A last disadvantage of the time domain is that an implementation of advanced correction or signal processing algorithms is often done easier in the frequency domain. While in principle it is possible to replace the frequency domain's simple complex multiplications by time domain convolutions with FIR or polyphase filters, in practice this often leads to more complicated computational solutions.

4   APPLICATION EXAMPLES

4.1   Plasma ignition field

In Figure 2, we show the time domain signal of a plasma ignition field as used in modern electronic combustion control systems for automotive applications. This plasma field is demonstrated as an example of a technical object emitting acoustic signals exhibiting extremely short pulses. While the time distance between the pulses themselves might still be sufficiently long enough to try applying a frequency domain approach, an individual pulse has a length of less than 100 microseconds which prohibits windowing and averaging.
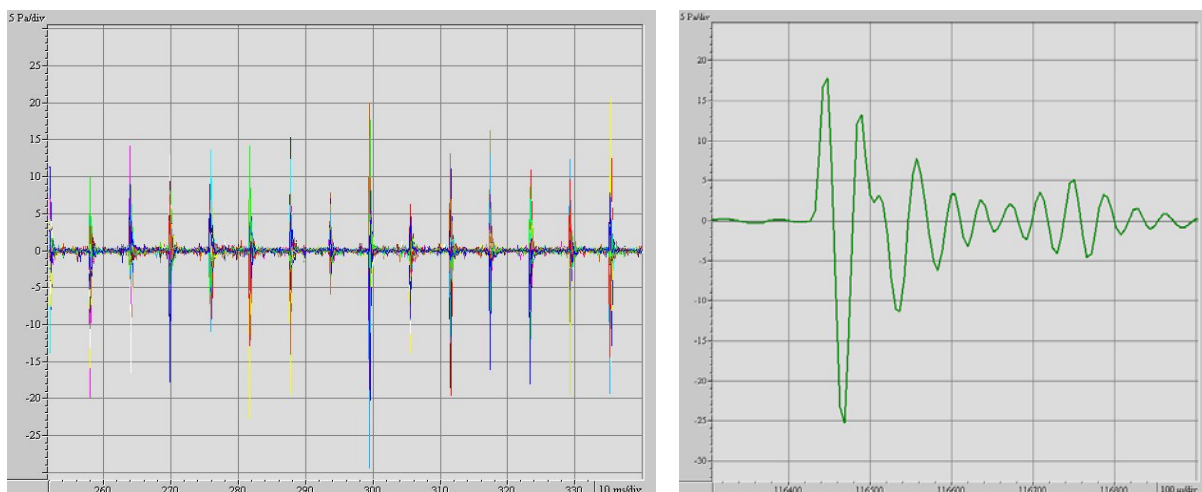
Fig. 2. Consecutive pulses of a plasma ignition field. Time spacing between pulses is 6 ms (left), but the duration of a single pulse is less than 100μs (right). The following peaks in the right picture are already reflections at the mounting plate of the individual sparking plugs.

In Fig. 3, the beamforming result of our time domain calculation is shown as an acoustic image showing a single ignition plug sparking. The array used was our standard ring array with 32 microphones and with a diameter of 72,5 cm. No signal prefiltering was applied here. The measurement distance was 66 cm, and the sampling rate for every microphone channel was 192.000 samples per second. The image contrast is 6 dB. The measurements were performed at a booth inside an exhibition hall without special acoustic shieldings.
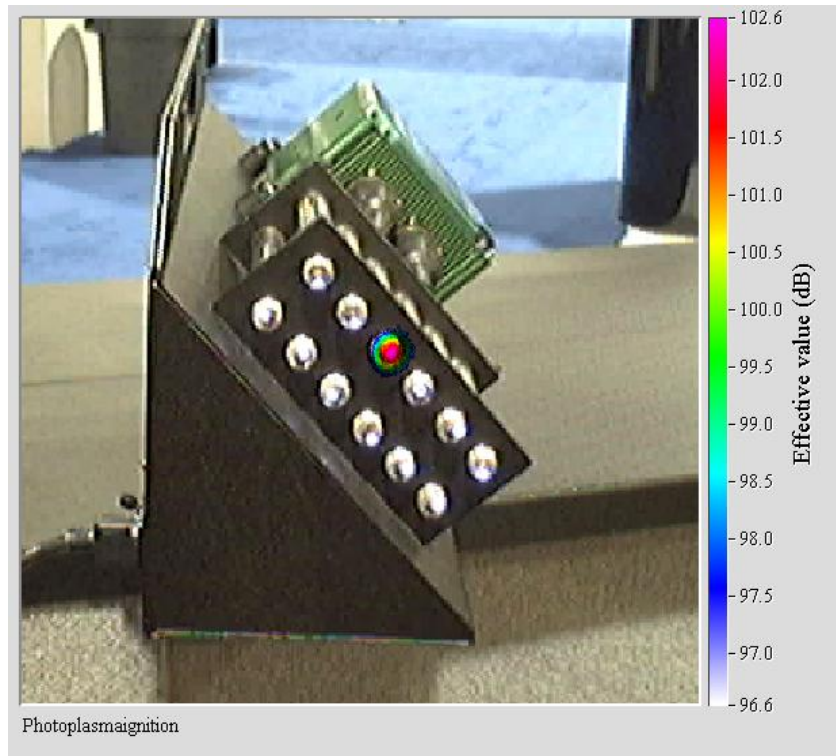
Fig. 3. Localization of an individual pulse of a plasma ignition field (Woodward company) with the Acoustic Camera in the time domain. Image contrast is set to 6 dB.

4.2   Nontechnical example - localization of a bat

In Fig. 4, an acoustic photo of a bat starting its flight from the wall is shown as an example of a somewhat exotic non-technical application. The signal was bandpass-filtered between 25 kHz and 50 kHz before performing the time domain beamforming calculation. The measurement distance was about 4 m, the array used and the sampling frequency were the same as in the technical example above. The image contrast is set to 9dB. This example demonstrates that even though high frequency aliasing effects are present it is easily possible to detect the source that is still appearing as the loudest emission. With broadband calculations, it is well possible to undersample in the spatial domain when there is enough time domain oversampling and an array geometry which is not too bad in contrast.
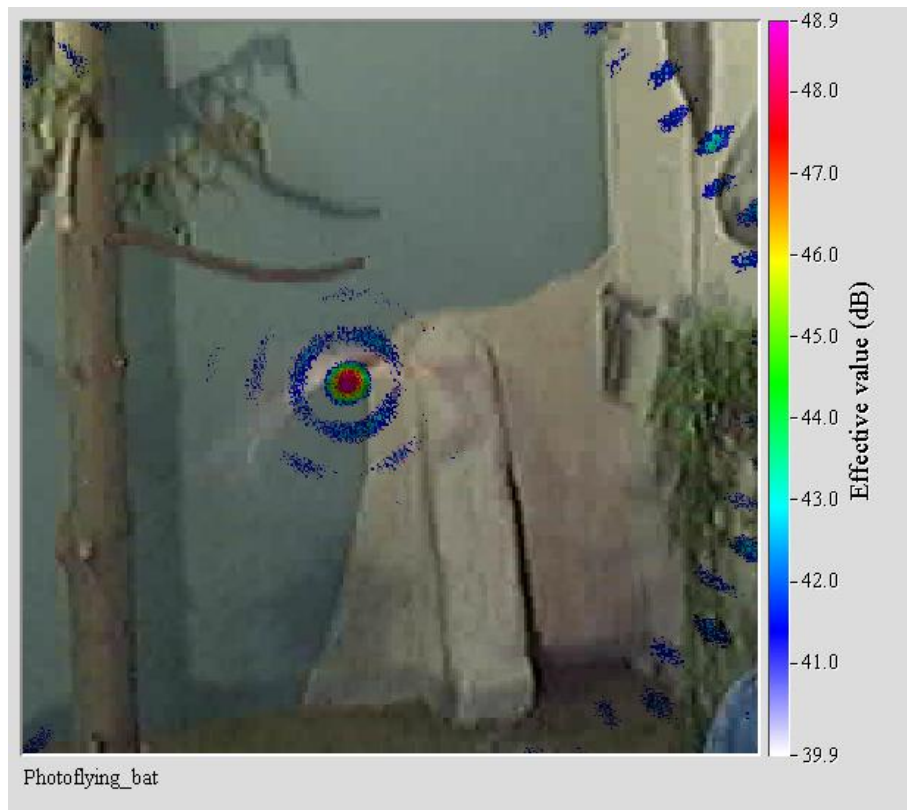
Fig. 4. Acoustic photo of a flying bat. Contrast here is 9 dB and clearly high frequency-aliasing is visible, but localization of the bat is possible without problems.

5   CONCLUSIONS

The paper compared the pros and cons of calculating the simple delay-and-sum beamforming in time domain and in frequency domain, pointing out some benefits of the time domain resulting from our practical measurement experiences. Among those benefits of the time domain approach are its inherent broadband behaviour leading to less aliasing problems, a relatively high computational efficiency and a very good applicability for non-stationary and strongly transient signals. Especially for such short-time events the simple time domain formula is often much better suited than the frequency domain approach.

Limitations of the time domain beamforming exist due to runtime delay quantization effects causing a need for high sampling rates and processing huge amounts of data. Also, many of the more advanced signal processing algorithms are easier to implement in the spectral domain.

Other well-known limitations as the bad contrast for low frequencies are not dependent on the calculation domain but rather have to be attributed to the physics of the beamforming method itself. Those effects are very complex and can only to be seen in the whole context of object size and distance, array geometry and aperture, image field and signal frequency content.

REFERENCES

[1]   Johnson, D.H.; Dudgeon, D.E.: *Array Signal Processing. Concepts and Techniques.*
       Prentice Hall PTR: Englewood Cliffs, NJ, 1993.

[2]   H. van Trees: *Optimum ArrayProcessing.* J. Wiley & Sons, 2002.

[3]   Naidu, P.S.: *Sensor Array Signal Processing.* CRC Press LLC, 2001.

[4]   Michel, U.; Barsikow, B.; Haverich, B.; Schüttpelz, M.:
       Investigation of airframe and jet noise in high-speed flight with a microphone array.
       Proc. of 3rd AIAA/CEAS Aeroacoustics Conf., Paper No. 97-1596, Atlanta, 1997.