Sound source localization (SSL) and quantification techniques using microphone arrays are now standard tools in aeroacoustic testing. These array systems give a rich view of the source distribution and strength over the tested object from an outflow point of view. But in the case of additional transfer to the final point of interest, for instance the ears of the driver in a car or of the pilot in an aircraft, it is useful to modify the sound source Localization algorithms in order to exhibit the important sources at the control point.

First the use of a reference sensor is presented, like the conditioned spectral analysis (CSA) to filter the cross spectral matrix (CSM). Some important parameters are analysed, that must be considered to have a correct evaluation of the constrained CSM.

Second, a modified Clean-SC algorithm is proposed. By conditioning the cross spectral matrix with a reference sensor, it is possible to assess the source strength relative to this reference sensor. Some further modifications of the algorithm are presented to keep both advantages of classical Clean-SC (the ability to deal with coherent sources) and of sensor conditioning. Discussions compared to the existing literature concerning this topic is given. This contribution presents details of the algorithm, its validation in a wind tunnel and applications on a car immersed in a flow.

Keywords: Clean-SC, coherence, beamforming, wind tunnel, aeroacoustics

1. Introduction

The combination of acoustic phased arrays and experimental wind tunnel set-ups has proved a powerful tool in the last decades to study the nature of noise sources of moving vehicles. In particular car manufacturers make an extensive use of such tools [1] since they allow first to exhibit the localisation of main aeroacoustic sources on a model under development. Microphone arrays help to point out parts of the vehicle to work on, with the highest priority. Second, they allow to estimate the acoustic power emanating from various specific regions, even if they are weak contributors to the far field noise. Such a paradoxical capability is of prime interest for cars in wind tunnels. For instance,
the side mirror is typically one of the most important noise sources from the driver’s point of view especially for the coincident frequency of the glass panel where the sound radiation into a cabin is efficient. However, there are usually other noise sources that are louder than the side mirror sources from the exterior point of view. The wheel noise is for example one of such a broadband noise caused by turbulence emanating from the interaction of the flow with the front wheel. The ability to emphasise noise sources that have a relevant contribution to the interior cabin using an external array is thus a very important matter. Another way to study cabin-significant sources is to place a 3D spherical array inside the car to draw 3D localisation maps [2]. However, the available space in the car cabin is limited and the view angle is generally not optimal to separate wheel noise and mirror noise. Conversely, external arrays are generally much larger than interior array and their position is ideal to distinguish the various sources on the car surface, especially in the lower frequency range. As a consequence, it is often wanted to estimate the sound power emitted from the mirror using the combined advantages of external and internal measurements. More precisely it is expected to draw source maps on the car surface from the external array data but with source amplitude values weighted by their contribution to the interior noise. Typically a binaural microphone signal or a beamformed time signal using an internal array provides the interior noise information. The part of the signal that is not coherent with the reference signal can be filtered out during the computation of the external microphone cross-spectra. Subsequently the conditioned cross-spectral matrix (CSM) is applied in the standard beamforming process to make a bridge between the interior and exterior noise. This beamforming approach using a reference sensor is meantime recognized as a standard analysis tool in wind tunnel tests.

The limitations of the beamforming algorithm have been identified for several decades now since maps are tarnished with spurious sidelobes that significantly limit the dynamic range of the method. The spatial resolution is also limited by the size of the array and the distance to the sources. In the lower frequency range, it makes difficult to separate close sources (as for instance the side mirror and the A-pillar corner). To overcome these limitations numerous advanced methods have been proposed in the literature. One of the most popular in the field of aeroacoustic is the Clean-SC method introduced by Sijstma [3]. Compared to classical beamforming, this technique provides a significant extension of the dynamic range and a slightly better separation power within a very short evaluation time. Moreover, source extension and acoustic reflection or scattering are accounted for, in the sense that all the energy that is coherent with a beamformed signal is attributed to one single point in the clean map, even if the actual sound propagation from the noise sources to the array microphones do not match to the steering vector of the beamforming. The combination of the acknowledged advantages of the Clean-SC method with the sensor conditioning methodology is the main topic of this paper. It is aimed at getting a similar dynamic range and separation capacity than the original formulation, while displaying the source amplitudes that are meaningful for interior noise study. Several formulations have been proposed in the literature to this end [5, 6], but it will be shown from simulations that they suffer from various limitations that motivated the present work. Our proposal on this paper is to apply a specified CSM in order to be able to map the coherence coefficient between exterior sources and a reference sensor. These values on the source map lead directly to the quantitative estimation of their contribution to the reference sensor.

In section 2 the formulation of the CSM and the conditioned-CSM including for the coherence map are described. Additionally, the impact of the statistical convergence on the CSM and conditioned-CSM is analysed. In section 3 the classical Clean-SC algorithm is briefly presented, and the sensor conditioned Clean-SC is introduced with comparison to the cited reference. In section 4, the validation of the proposed approach is shown by performing simple simulations and wind tunnel experiments.
2. Conditioned conventional beamforming

2.1 Formulation

The first part of this section presents the matrix formulation of conventional beamforming in the frequency domain, as it is the ground of subsequent tools. The cross-spectrum between two sensors \( i \) and \( j \) at frequency \( f \) is classically given by

\[
S_{ij}(f) = \mathbb{E}\{s_i(f)s_j^*(f)\} 
\]

Where \( s_i(f) \) is the Fourier component of the signal of sensor \( i \) and \( \mathbb{E} \) is the expectancy operator evaluated by the ensemble average over time snapshots. Considering \( N_m \) array sensors, the cross-spectra between all sensors pairs can be organized in the \([N_m, N_m]\) cross-spectral matrix \( S(f) \) whose entries are \( S_{ij}(f) \). Scanning a specific grid point with Conventional Beamforming (CB) consists in using a steering vector \( h(f) \) of dimension \([N_m, 1]\) and computing the beamformed output

\[
b(f) = \frac{h^*(f)S(f)h(f)}{|h^*(f)h(f)|^2} 
\]

Where \( h^* \) stands for the conjugate transpose operator. Vector \( h \) is typically filled with a model of the propagation transfer function at frequency \( f \) between the scan point and each microphone. In the specific case where a single source is present, actually located at the scan point, and that the steering vector corresponds to the true propagation vector, the quantity \( b(f) \) provides a good estimate of the source squared amplitude even in the presence of noise.

In the following we are presenting the sensor conditioned beamforming. Assuming that a reference sensor \( r \) is available and that the corresponding time series is synchronized with the phased array, it is possible to compute a filtered cross-spectrum \( S_{ij;r}(f) \) between sensor \( i \) and \( j \) conditioned by sensor \( r \).

\[
S_{ij;r}(f) = \frac{s_{ij}(f)s_{r}(f)}{s_0}, \quad S_0 = S_{rr}(f) 
\]

The numerator of \( S_{ij;r}(f) \) represents the energy in the cross-spectrum \( S_{ij}(f) \) which is correlated with the signal of the reference sensor \( r \). Note that this numerator of Eq. (3) shows only the crossspectra involving the reference sensor (and not directly the cross-spectrum between array channel \( S_{ij}(f) \)). The expression of \( S_{ij;r}(f) \) is given typically normalised by the autospectra of the reference sensor \( S_{rr}(f) \). In a very similar way as previously, the conditioned CSM \( S_r(f) \) can be composed of \( S_{ij;r}(f) \). The beamforming output \( b_r(f) \) filtered by the reference sensor can be calculated by applying \( S_r(f) \) instead of \( S(f) \) in Eq. (2).

\[
b_r(f) = \frac{h^*(f)S_{r}(f)h(f)}{|h^*(f)h(f)|^2} 
\]

As long as the CSM \( S_r(f) \) contains the numerator of \( S_{ij;r}(f) \), the beamforming output \( b_r(f) \) provides always in a sense of a correlated source map to the reference sensor. The physical meaning of \( b_r(f) \) is, however, depending on the selection of \( S_0 \) in Eq. (3). As we are interested in a quantitative relationship between the level of exterior sources and a reference sensor, it makes sense to determine the coherence value of them. In order to calculate the beamforming output as a coherence value, \( S_0 \) to be defined by following:

\[
S_0 = \sqrt{S_{ii}(f)S_{jj}(f)S_{rr}(f)} 
\]
Applying the Eq. (5) into the Eq. (3), finally the coherence coefficient map can be calculated by Eq. (4). This coefficient might be expressed as a contribution coefficient from exterior sources to the reference sensor. In an ideal case, if uncorrelated sources are found and cleanly distributed on a source map without any lobes, so that the power of the reference signal is completely composed by only those exterior sources, the integration over the coherence map becomes 1. One of the reasons why in practice this can not happen is due to the convergence issue on the cross-spectrums which will be described in the next subsection.

### 2.2 Convergence

Evaluated cross-spectra in equation 1 are tarnished with measurement noise, which is linked to the quality of the convergence of the ensemble average operator. From equation (2), it can be seen that this error is directly propagated to beamforming output, although mitigated through the beneficial effect of the summation over all microphones. To produce proper beamforming maps, one should ensure that cross-spectra are averaged over a sufficient number of snapshots such that the statistical noise lies well below the beamforming dynamic lower bound. In conventional beamforming applications with large arrays (i.e. if the number of microphones can reach a hundred), a few seconds of recording are sufficient to ensure satisfactory convergence. However in the coherent beamforming formulation (eq. 4), the number of ensemble average should be significantly increased which leads to require longer recorded time signals. This effect is studied here below.

From Bendat and Piersol [4], the normalized random-error for the cross-spectrum evaluation is given by:

$$\varepsilon[|S_{ij}(f)|] = \frac{1}{|\gamma_{ij}(f)|^{\frac{1}{2n}}}. \quad (6)$$

Where \(n\) is the number of the ensemble average and \(\gamma_{ij}(f)\) is the square root of the coherence function between sensors \(i\) and \(j\). The first remark is that \(\varepsilon\) is inversely proportional to \(\sqrt{n}\), which means that it becomes more and more difficult to reach the required small error value by increasing \(n\). The second remark is that \(\varepsilon\) is inversely proportional to \(\gamma_{ij}(f)\). For microphone signal with low coherence, in order to get a low error value \(\varepsilon\), therefore, a higher number of ensemble average is needed in compared to those of highly correlated signals. This is not noticeable for CB since microphone signal from the same array are usually highly correlated.

For the correlated beamforming, the normalized random-error for \(S_{ij,r}(f)\) can be expanded as following:

$$\varepsilon[|S_{ij,r}(f)|] = \frac{1}{\sqrt{n}}\left(\frac{1}{|\gamma_{ir}(f)|} + \frac{1}{|\gamma_{jr}(f)|} - 1\right). \quad (7)$$

The meaningful quantity here is the square root of the coherence \(\gamma_{ir}(f)\) between an array sensor and the reference sensor. In case the reference sensor is located in the car cabin, the coherence with an array microphone is generally poor when dealing with wind excited noise: \(|\gamma_{ir}(f)| \ll |\gamma_{ij}(f)|\). This implies an increase in the error noise. Moreover since \(|\gamma_{ir}(f)| \approx |\gamma_{jr}(f)|\), we have \(\varepsilon[|S_{ij,r}(f)|] \approx \frac{1}{\sqrt{n}}\left(\frac{2}{|\gamma_{ir}(f)|} - 1\right)\); the factor 2 in the coherence dependence generates a further increase of the error noise, which must be compensated by the number of ensemble average \(n\). Therefore, the acquisition time length must be chosen carefully especially for the conditioned-CSM, in order to reduce the error value \(\varepsilon\) depending on frequencies and the general coherence condition between array sensors and the reference sensor. The above explained the relationships between \(\varepsilon\) and \(\gamma\) are plotted by choosing different \(n\) in the following.
In a practical example case, when the coherence between the reference sensor and the microphones of the array is about 0.1, the ensemble average should be at least carried out 2835 times in order to get the relative error of 10%, which corresponds to the time signal length of 141 seconds for a frequency resolution of 20 Hz. In the case that only 30 or even 10 seconds of signals are recorded, the relative error would become 21% and 37% respectively.

3. Conditioned Clean-SC

3.1 Classical Clean-SC

In this section, the standard Clean-SC method is briefly presented for completeness. Starting with a null clean map, the iterative procedure is the following [3]:

1. Obtain a source plot with classical beamforming from eq. 2 (dirty map)
2. Clean map update:
   a. Search for maximum peak location in the dirty map \( \xi_{max}^{(i)} \) with power \( P_{max}^{(i)} \)
   b. Store power \( P_{max}^{(i)} \) in the clean map at location \( \xi_{max}^{(i)} \)
3. CSM cleaning:
   a. Compute the coherent steering vector \( h^{(i)} \) using the (degraded) CSM and the focused signal in \( e_{max}^{(i)} \)
   b. Compute the CSM induced by the source in \( \xi_{max}^{(i)} \cdot G^{(i)} = P_{max}^{(i)} h^{(i)} h^{H(i)} \)
   c. Compute the degraded CSM \( S^{(i)} = S^{(i-1)} - G^{(i)} \).
4. Loop to step 1 for iteration (i+1) until the degraded CSM energy stops decreasing.

3.2 Conditioned Clean-SC

The Clean-SC algorithm is modified to take into account the information of the reference sensor. The modification mainly concerns a parallel use of the original CSM \( S(f) \) and the conditioned CSM \( S_{r}(f) \). Using the same notations, the proposed procedure is the following:

1. Compute the conditioned CSM \( S_{r}(f) \) with reference sensor \( r \) from equation (3). Note that this CSM contains also the reference sensor.
2. Clean map update:
   a. Obtain a source plot with coherent beamforming using equation (4) (conditioned dirty map) and \( S_{r}(f) \)
   b. Search for maximum peak location in the conditioned dirty map \( \xi_{max,r}^{(i)} \) with power \( P_{max,r}^{(i)} \) and store it in the clean map.
3. CSM cleaning:
   a. Using the original CSM \( S(f) \), compute the coherent steering vector \( h^{(i)} \) in \( \xi_{max,r}^{(i)} \).
   b. Compute also the power \( P_{max}^{(i)} \) without the reference sensor at the location of the referenced map maximum, i.e. evaluate equation 2 using \( S(f) \) at \( \xi_{max,r}^{(i)} \).
   c. Compute the CSM induced by the source in \( \xi_{max,r}^{(i)} : G^{(i)} = P_{max}^{(i)} h^{(i)} h^{H(i)} \)
   d. Compute the degraded CSM \( S^{(i)} = S^{(i-1)} - G^{(i)} \), also remove coherent part of \( P_{Max} \) on the cross spectrums containing the reference sensor.
   e. Compute the degraded referenced CSM \( S^{(i)}_{r} \)
4. Loop to step 2 for iteration (i+1) until the degraded CSM energy stops decreasing.
3.3 Comparison with previous publications


At first Adam [5] has use the conditioned CSM to compute the beamforming, then the Clean PSF is applied. The limitations of this approach are the same than clean PSF. If the sources are spatially extended or if the propagation is not like assumed for PSF generation, the results are degraded. Furthermore, this deconvolution is heavy in terms of computation.

For Hald [6], the Clean-SC algorithm is modified to compute a contribution (or correlation) coefficient at each loop of the algorithm. In other words, each monopoles identified in the Clean-SC deconvolution are used to process the ratio between conditioned and standard beamforming. This approach is close to the one presented here except that it presents the same issues than Clean-SC for low frequencies. If the sources are not separated by the beamforming algorithm, the maximum position is wrong. This behaviour is improved by the method presented in this paper.

4. Validation of the proposed method

4.1 Simulation using two uncorrelated point sources

As a first validation case a sound field generated by two uncorrelated point sources is studied. These two sources are located in a same plane, marked as a cross in Figure 2 and 3, with an equal power density. The reference signal contains the exact same signal as the right side source for this study. The array aperture and the number of the microphone are designed such that the two sources cannot be separated with the beamforming source mapping for frequencies below 1 kHz. The coherence between the two sources was plotted, the values are around $10^{-3}$ and it can be assumed that those two sources are uncorrelated to each other.

First of all, Figure 2 shows the Clean-SC output for different $1/3$rd octave band frequency. In order to be able to check the localisation points and the source strength precisely, no interpolation is used in these maps.

As designed, the standard source mapping can neither separate the two sources nor correctly for the given location until 1250 Hz. This designed array performance makes Clean-SC to find its maximum in-between the two sources which leads to a wrong interpretation that there is either only one source, or more than two sources at wrong locations for those frequency bands.

![Figure 2: Clean-SC applied on two uncorrelated monopoles for different $1/3$ octave bands.](image)

The Coherence coefficient applied with Clean-SC is evaluated using Eq. 4) based on the algorithm described in 3.2 and shown in Figure 3. In comparison to Figure 2 the right side source can be already located correctly by considering the reference sensor signal at 630 Hz.
In terms of the coherence coefficient, the level on the right source is around 60% for 630 Hz. The right side of the coherence gets close to around 99% while the left side of the coherence points vanish if frequency band increases. This is caused by the low spatial resolution of the beamforming. For example, at 630 Hz, the beamforming algorithm locates the source in the middle of the two sources that is a sort of the mixing of two sources. The correlated beamforming enables to locate the right side source correctly but with a wrong source strength. As a consequence, this small spatial mismatch cause an underestimation of the coherence coefficient. As soon as the beamforming resolution is high enough to separate the sources, the coefficient is correctly estimated.

4.2 Wind tunnel tests

The experimental tests for the proposed beamforming approach was carried out at the Daimler Wind Tunnel Centre in Sindelfingen. Using the current C-class Coupe the wind noise performance was evaluated at the wind speed of 140 km/h (0° yaw angle). First of all, the radiated exterior sound power is recorded at the microphone array system which is located 6m from the ground. Additionally, the interior noise was measured using binaural heads synchronized with the array microphones together in order to make the sensor conditioned beamforming possible. The right ear of the front passenger’s head was used as a reference sensor for the study.

Figure 4 shows the beamforming results for the one third octave band at 2 kHz in depending on different time duration of recorded signals from 10 seconds to 300 seconds. The convergence behavior of the source map can be analyzed by changing only the number of the ensemble average for the determination of the cross spectral matrix. All maps are giving the same levels on every single sources and the source distribution on map does not change among those 4 different signal lengths which means that the convergence of cross spectrum is achieved with a very short signal length (with 10s recording time, is already sufficient).
Figure 5 presents the same convergence study on the coherence coefficient depending on different signal length. In this case, since the absolute coherence between array microphones and the reference sensor is low, the required time for the convergence is long. By comparing the different amount of the ensemble average it is clear that the evaluation using 10 seconds of recorded signals is, at least at this 1/3rd octave frequency, not good enough to make a correct interpretation. The sources at the vehicle front area as well as at the right side of the A-pillar bottom area, which are visible at Fig. 5 a) with relatively higher coherence level, will be suppressed by increasing the number of the ensemble average. Those sources have at the end no such a relevant contribution to the reference sensor (the right ear of the front passenger). One would have made a wrong interpretation if the evaluation would have been performed by using only 10s of test signal.

Figure 5. Coherence for 2kHz 1/3rd octave band, using different signal length: 

a) 10 s, b) 30 s, c) 60 s, d) 300 s

Figure 6 shows the standard beamforming map a) and the coherence coefficient to the reference sensor b) as well as the Clean-SC evaluation of them c) and d) using 60s recorded time signal. Although the standard source map already gives the general idea about the localization of noise sources, the cleaned map can indeed show us a precise spatial information of the sources and the hidden sources under the side lobes. One should be aware of that the source maps were, in this study, evaluated on a certain 2D plane at the constant distance from the ground where the side mirrors are located. Due to this 2D plane evaluation, the sources which are not locating on this plane were forced to be projected on this evaluation plane. Those sources cannot be therefore displayed at the correct position on the evaluated plane, which is partly visible especially in c). The beamforming with the coherence coefficient provides extreme valuable information since the all detected exterior noise sources can be in an ideal case sorted after the quantity of each source’s contribution to the reference sensor. Figure 6 b) and d) are therefore highlighting the exterior noise which are relevant to, in this study, the right ear of the passenger’s head. Using this proposed approach, the aeroacoustics performance in terms of the inside of a vehicle cabin can be optimized very efficiently.
5. Conclusions

The conditioned-CSM formulation was presented for the coherence beamforming mapping using a reference sensor. As the mapped coherence corresponds to the measure of each sources contribution to the reference sensor, the detected exterior noise sources can be efficiently attacked in order to reduce the sound pressure at the reference sensor. Additionally, the combination of the Clean-SC to the conditioned-CSM was demonstrated successfully. The conditioned-CSM needs however, significantly higher number of the ensemble of average in comparison to the conventional CSM in order to get same order of the error for the CSM calculation, if the coherence between microphones of the array and the reference sensor is particularly low.

REFERENCES