



# EFFICIENT METHODS FOR THE ANALYSIS OF MOVING SOUND SOURCES WITH HIGH CHANNEL MICROPHONE ARRAYS

Andy Meyer and Dirk Döbler  
GFaI e.V.

Volmerstraße 3, 12489, Berlin, Germany

Marcus Gräser  
gfaitech GmbH

Volmerstraße 3, 12489, Berlin, Germany

## ABSTRACT

The localization of stationary and transient sound sources on moving or static objects by multichannel arrays is State of the art for several years. In many industrial sectors (e.g. automotive industry, aircraft industry...) this method is indispensable for the noise reduction, noise analysis and error detection. Analysing moving sound sources increased requirements of hardware and software components. With the aid of light barriers or GPS in combination with inertial navigation systems (INS), it is possible to determine the mostly unknown average speed, or to estimate the speed as a function of time. Also entry and exit times of the moving objects can be determined. For smaller measurement campaigns, this method is not efficient.

For fast moving, long objects (e.g. rail vehicle, truck) often it is not possible to take a photo of the entire object with the built-in optical camera during the measurement. Hence, usually a draft is overlaid with the acoustical image manually, which, however, often does not match the actual object completely and therefore the exact localization of the acoustic sources proves difficult.

Furthermore, a frequency shift due to the Doppler effect is present in the sound signals affecting the runtime and phase information which are essential to Beamforming. This leads to significant spatial and spectral blurring of sound sources in the acoustical image.

Efficient methods are presented in this paper which allow the solution of solving the complex tasks of generating optical photos of the entire object, overlaying the images automatically and determining the velocity profile as well as enter and exit times using a microphone array with an integrated camera. Another section deals with the algorithm for the determination of the source position under consideration of the Doppler effect.

## 1 INTRODUCTION

Mobility in our society has become an increasing necessity. The further development of globalization creates greater demands on the air, rail, sea and road traffic. This is unfortunately accompanied by increasing human effects on the environment (e.g. noise, air pollution...). About a quarter of the population of the EU (according to the EU Commission) and about 60 % of the German population (according to the Federal Environment Agency) are directly affected by traffic noise. The negative effects of noise on humans are scientifically proven long ago (<http://www.fv-leiserverkehr.de/>).

On the initiative of the DLR (German Aerospace Center), since 1999 many partners from industry, authorities and research have been devoted to a major national research network (“Leiser Verkehr”). Within this network there is a project “Sound source localization”, which deals with the validation of acoustical simulation methods. This network shows that the localization of sound sources plays an increasingly important role.

Beamforming in combination with advanced algorithms as a method making sound and thus noise sources visible, has been used for over two decades as a tool to analyze sound sources and thus to reduce noise [1] in the field of railway vehicles. In the analysis of vehicle models [4] (for example 1:25) in a closed wind tunnel and in real pass-by measurements this method provides good results.

Pass by measurements still represent a particular challenge, because in addition to the required Dedopplerization of the signals, the following parameters must be determined by additional parallel measurements:

- Complete optical image of the object
- Entry and exit times of the object in the measuring section
- Speed profile of the object in the measuring section for the Dedopplerization

The paper uses the example of a high speed train passing by, to show how the videos from the built-in optical camera reduces the time for preparation, measurement and evaluation through automated collection and calculation of the above mentioned parameters. Another advantage is the saving of any additional measurement hardware (Light barrier, GPS, etc.).

## 2 REQUIREMENTS FOR THE MEASUREMENT SETUP

A standard technical configuration for this kind of measurements is multichannel microphone arrays with integrated optical cameras. The expected lower and upper cut-off frequencies determine the dimension and the number of microphones of the array.



*Fig. 1 : Measurement setup with a 120 channel microphone array (spiral) with built-in industry camera*

In order to achieve the required video quality of the high-speed train, it is essential to select a video camera with sufficient frame rate (at least 60 frames per second) and short exposure time (1/1000 seconds per image). At a speed of 300 km / h (83.3 m / s) the object is moving at least 1.3m per frame at a frame rate of 60 FPS (60 FPS  $\triangleq$  16 ms per frame). Precondition for such a small exposure time is sufficient lighting of the measured object.

### 3 PANORAMIC IMAGE FROM OPTICAL FLOW

At a measuring distance of about 5 to 10m from the passing by object (train, tram, truck...), it is not possible to completely capture the object in a single frame with an optical camera. If you create a picture with an additional camera from a larger distance, this picture will not be very detailed and will have perspective distortion at the edges. Furthermore, there would be a lack of geometric reference to the microphone array and the built- in optical camera. So far a solution to the problem has been that a line draft of the object was superimposed by an operator using the acoustic image. This highly subjective sort of combination was done with the aid of reference sources (e.g. pantograph of an ICE 3). For completely unknown objects (diesel locomotive with wagons) mappings of the acoustic emissions to the relevant object parts were extremely difficult.

One possible solution was presented in [2]. While the vehicle is traveling at a velocity ( $V$ ) passing the video camera, this camera records a number of  $N$  images (frames)  $F_n (n = 1 \dots N)$ . In two subsequent frames  $F_n$  and  $F_{n+1}$  unique and corresponding reference points (feature points) are searched for in the optical flow part of the picture, to calculate the translation parameters between these two images. With these parameters a panoramic image containing the complete vehicle can be created successively.

Unfortunately, the developed algorithm still has crucial weaknesses, which are also described in [2].

1. The video camera must be perfectly aligned parallel to the passing by object (e.g. to the rails).
2. The speed of the passing by object must be constant.

Furthermore, there are areas in the case of passing by trains that might make an automatic assignment of reference points very prone to errors due to recurring segments (windows, wheels...).

To overcome these weaknesses, a fully automatic algorithm has been developed which makes it possible to extract the moved parts of two images  $F_n$  and  $F_{n+1}$  from a video and using a correlational method, to identify the offset between the two images in X and Y - direction.

This will be explained in the following.

#### 3.1 Correlative algorithm for generating a panoramic image

To reduce the computation time, the optical flow is segmented in the first step. There are many different ways in literature (e.g. segmentation by color gamut, edge detection or model - based methods). In the present case, a threshold method with a prior differential image calculation was implemented. The advantage in this robust approach is that the effect of image defects such as noise, brightness differences and other moving objects (trees) is minimized on the image segmentation.



Fig. 2 : two consecutive video images of a passing ICE at about 280 km / h

Without the use of the threshold method it very often happens that parts of an image are segmented which have not moved (Fig. 3 left).

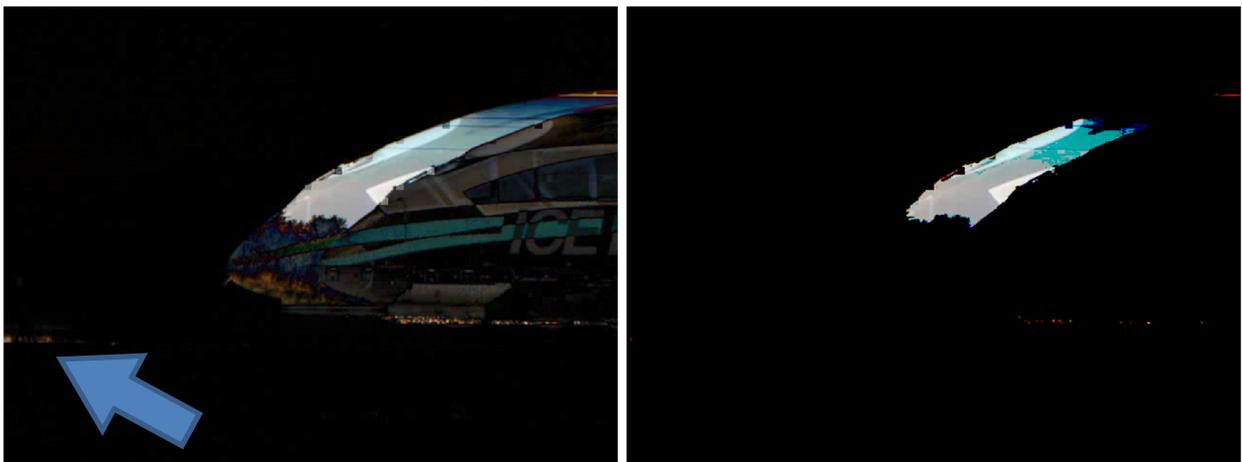


Fig. 3 : difference image of the two images of Figure 2 without threshold (left) and with threshold (right)

By introducing a suitable threshold value, these error influences could be minimized (Fig. 3 - right) with the consequence that even moving parts are no longer segmented. This circumstance could be considered acceptable.

After the separation of still and moving image parts, the correlations of the moving image section of  $F_n$  (marked red in Fig. 4) with the entire image  $F_{n+1}$  over the three color channels can be calculated. For this purpose, the cross-correlation matrix  $r_{xy}$  by formula (1) was calculated where  $F_n(i)$  is the  $i$ -th intensity,  $\overline{F_n}$  the average of the intensities and  $\sigma(F_n)$  is the standard deviation of the intensities from the image  $F_n$ . The cross-correlation coefficients have their maximum in the displacement vector  $\vec{t} = (\Delta x, \Delta y)$  between two consecutive frames.

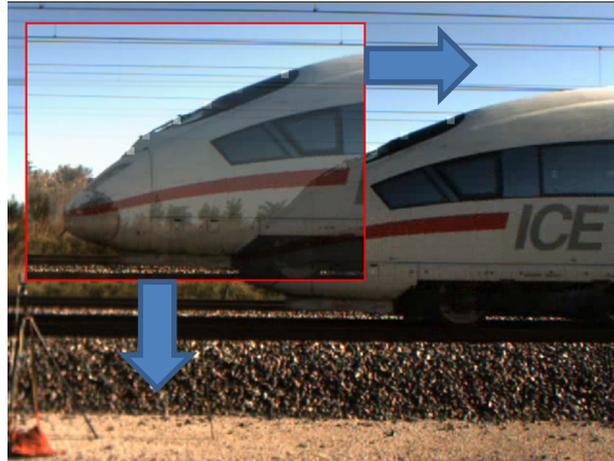


Fig. 4 : Principle of image correlation

$$r_{xy} = \frac{\text{Cov}(F_n, F_{n+1})}{\sigma(F_n)\sigma(F_{n+1})} = \frac{\sum_i ((F_n(i) - \bar{F}_n)(F_{n+1}(i) - \bar{F}_{n+1}))}{\sqrt{\sum_i (F_n(i) - \bar{F}_n)^2} \sqrt{\sum_i (F_{n+1}(i) - \bar{F}_{n+1})^2}} \quad (1)$$

These calculations result in  $N - 1$  translation vectors which form the basis to compose the panoramic image. Statistical tests can be used to detect and eliminate coarse errors.

### 3.2 Calculating the velocity profile

The basis for the consideration of the Doppler effect is a known velocity or a velocity profile (accelerating while passing). With the displacement vectors and the distance from the microphone array to the object, such a profile can be derived. The sign of the velocity ultimately provides the traveling direction of the measured object.

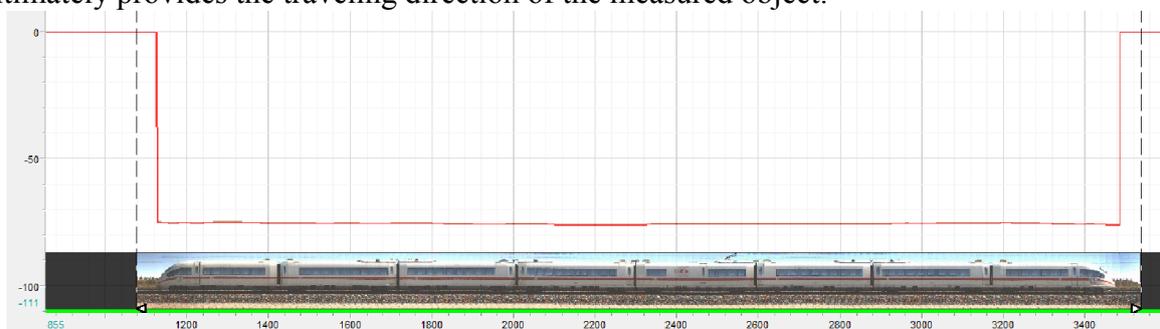


Fig. 5 : shows speed variation over time of an ICE

### 3.3 Compensate for the misalignment of the array

Usually it can be assumed that the array at the time of measurement is nearly horizontal and that the object from left to right or vice versa goes through the image. If the array is not sufficiently accurately aligned, this leads to a vertical offset in the panoramic image (Fig. 6).



Fig. 6: vertical offset by misalignment of the microphone array during measurement (overview)

Not accurately aligned means in this case, that the horizon of the optical camera is not parallel to the direction of the velocity vector of the object.

By an image transformation, this rotation can be compensated for (Fig. 7 and Fig. 8). The inclination of the array (microphone positions) must be compared with the panoramic image and corrected.



Fig. 7 : after compensation of the vertical offset (left detail)



Fig. 8 : after compensation of the vertical offset (right detail)

#### 4 THE ACOUSTIC ALGORITHM

A well usable algorithm was already described in [5]. For measurements on high-speed trains it has been shown that the width of the tracking window (steering range) has large impact on the (acoustic) image quality. For the mapping of low frequencies the selected window shouldn't be too small (Fig. 9), otherwise the reconstructed time function is too short (capped waves, poor averaging). For higher frequencies a too wide tracking window is unfavorable (Fig. 14) since then enter and exit positions are focused very oblique to the sources and then the sources smear away (unfavorable resolution of the array).

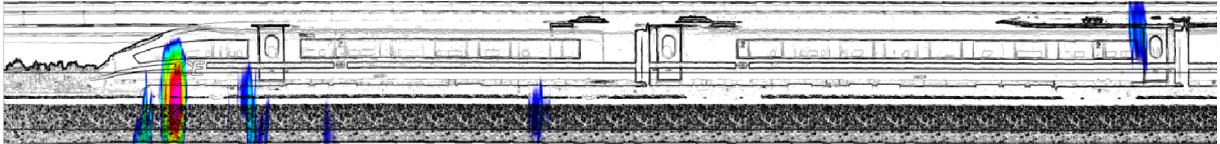


Fig. 9 : 3<sup>rd</sup> octave band 500 Hz, 1m steering range

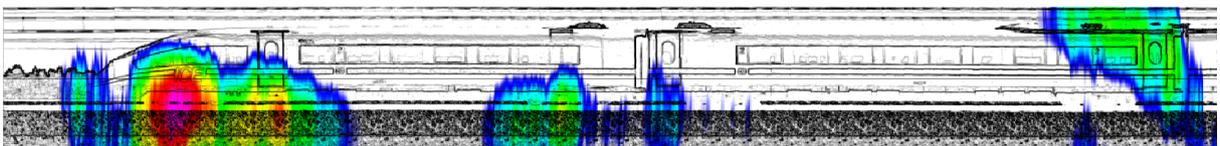


Fig. 10 : 3<sup>rd</sup> octave band 500 Hz, 5 m steering range

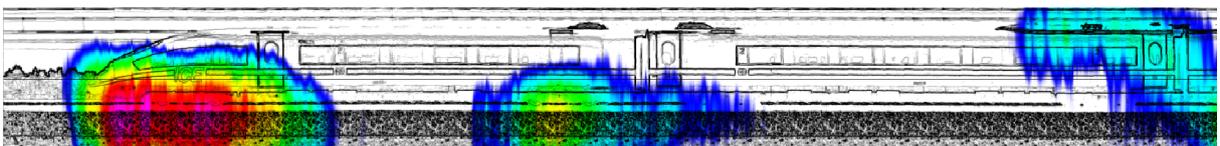


Fig. 11 : 3<sup>rd</sup> octave band 500 Hz, 10m steering range

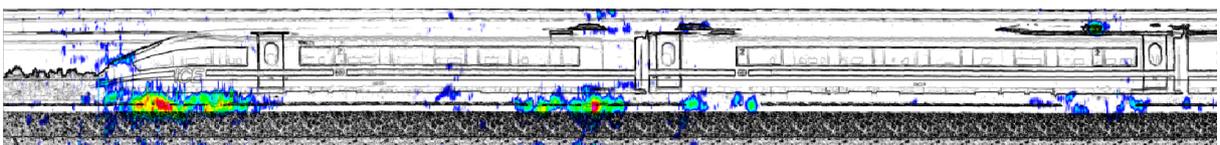


Fig. 12 : 3<sup>rd</sup> octave band 4000 Hz, 1m steering range

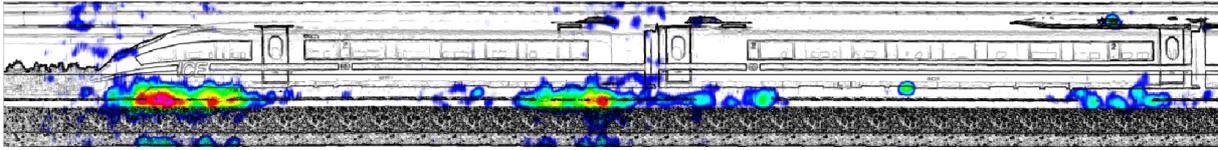


Fig. 13 : 3<sup>rd</sup> octave band 4000 Hz, 5m steering range

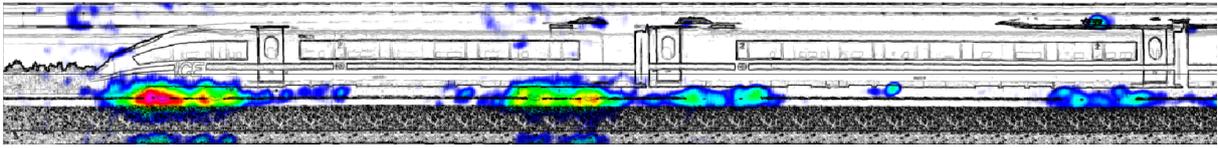


Fig. 14: 3<sup>rd</sup> octave band 4000 Hz, 10m steering range

Moreover, one can assume that many sources radiate perpendicular to the surface (especially rails and wheels), so that the array records these signals mainly when the sources are located in front of the array. Finally, the distance-dependent attenuation of the sound signals need to be considered. If it is already greater than the possible array contrast due to the distance from the array, the tracking window is chosen too large in any case. A proposal to solve these conflicting demands is the use of frequency-dependent tracking widths.

## 5 EXAMPLES

### 5.1 Tram

Fig. 15 shows a tram in central projective depiction (note the front area of the tram). The panoramic image was composed entirely from partial images which were copied from the right edge of the video image. The optical effect shown in Fig. 15 is a result of this procedure. However, in the acoustical image there are no perspective distortions, since the calculation is always referenced to a two-dimensional plane perpendicular to the main direction of the array.



Fig. 15 : panoramic image of a tram in central projection

An approximate parallel projection can be obtained by copying from the centre of the video image. The result of this procedure is depicted in Fig. 16. Now, the optical panoramic image and the acoustical photo match better.



Fig. 16: panoramic image of a tram in parallel projection

## 5.2 ICE

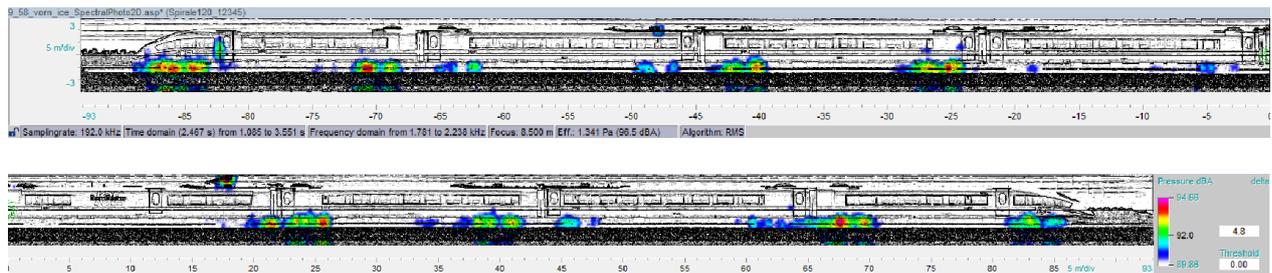


Fig. 17 : German high-speed train at about 270 km/h

## 6 CONCLUSIONS AND OUTLOOK

It has been shown that it is possible to analyse pass-by measurements fully automatically by evaluating the data of already available hardware e.g. a microphone array with integrated optical camera. Therefore, the following parameters are determined by a correlation algorithm from image information (video) and possibly compensated:

- Panoramic picture of the optical flow
- Enter- and exit-times of the moving object
- Velocity profile
- Alignment error of the array

Based on these parameters and applying a specially adapted time-domain Beamforming algorithm including Doppler correction, the acoustic map can be calculated without any further post processing. Thus, the complexity of the measurements as well as the analysis decreases significantly and the usability is raised.

As described in [6], different advanced algorithms should be investigated in order to improve the acoustical results. Also, the quality of the panoramic image can be increased by using a camera with a higher resolution. Eventually, refining the segmentation for the determination of the optical flow may lead to an improved robustness of the procedure.

## ACKNOWLEDGEMENT

This work has been funded by the German Federal Ministry of Economics and Technology (BMWi) under the project registration number MF100107.

## REFERENCES

- [1] Barsikow, B., King III, W., Pfizenmaier, E.: "Wheel/rail noise generated by a high-speed train investigated with a line array of microphones." *Journal of Sound and Vibration* 118(1),99-122 (1987)
- [2] Stier, J., Beitelshmidt, M., "Generating train side views from video sequences for microphone array PASS-BY measurements", 4<sup>th</sup> Berlin Beamforming Conference, Berlin, 2012
- [3] Clever, S., : „Ermittlung von Geschwindigkeitsverteilungen aus Farbbildern unter Verwendung des Korrelationsverfahrens“, Studienarbeit, Technische Universität Darmstadt, Institut für Automatisierungstechnik, 2004

- [4] Lauterbach, A., Ahlefeldt, T., Kröber, S., Koop, S., Loose, S.: „Microphone array measurements on high-speed trains in wind tunnels“, 3<sup>th</sup> Berlin Beamforming Conference, Berlin, 2010
- [5] Zechel, G., Zeibig, A., Beitelshmidt, M.: „Time-domain beamforming on moving objects with known trajectories“, 3<sup>th</sup> Berlin Beamforming Conference, Berlin, 2010
- [6] Michel, Dr. U., Barsikow, B.: „Localisation of sound sources on moving vehicles with microphone arrays“, Euronoise 2003, Naples, 19-21 May 2003