CONTRAST IMPROVEMENT AND SOUND RECONSTRUCTION OF QUIET SOUND SOURCES USING A HIGH DYNAMIC RANGE ALGORITHM

Dirk Döbler and Dr. Ralf Schröder
GFaI e.V.
Volmerstrasse 3 , 12489 Berlin, Germany

ABSTRACT

The maximum contrast of a standard beamforming result (acoustic photo) is characterized by the difference between mainlobes and sidelobes. Typical values for a standard beamforming system are 8 dB up to 15 dB, quieter sound sources are covered. Contrast improvement by simply increasing the number of microphones is expensive and ineffective. On the other hand, the increase in computational power allows the use of more complex algorithms to improve the acoustic map. Two years ago we presented at the BeBeC 2010 an algorithm to discover the masked sources, called “Acoustic eraser” [1]. This paper introduces a time domain beamforming algorithm based on the acoustic eraser algorithm. The algorithm provides an automatically, successively decomposition of the sound field with a high dynamic range. That allows contrast improvements up to 50 dB. Due to this high contrast quiet source can be discovered and a very accurate sound reconstruction is practicable.

1 INTRODUCTION

The low contrast of standard beamforming [2] is a limiting factor for the application of the technique. If the differences between the sound levels of sources in an acoustic scene are greater than the maximum contrast of the used microphone arrays (the level distance between main lobe and sidelobes), weaker sources are masked by the sidelobes of the stronger sources. Furthermore, crossing sidelobes from different sources can create “ghost sources”, a superposition and amplification of the crossing sidelobes. A characteristical example is illustrated in (Eq.1, Fig. 1). It shows an acoustic photo of five uncorrelated noise sources: two sources are visible, three sources are covered. To make the lower sources visible, in [3] and [4] main sources are eliminated by correlative methods that causes an improvement of the contrast of the beamforming result.

$$\boxed{\hat{f}(\vec{s}_1,t)=\frac{1}{M}\sum_{i=1}^{M}w_i y_i\left(t-\Delta t(\vec{s}_1)\right)} \qquad (1)$$

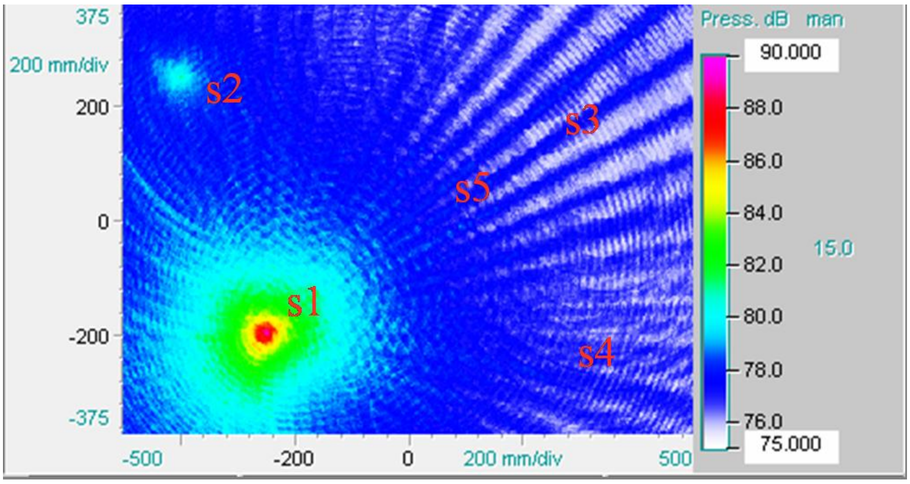

Fig.1: Acoustic photo of 5 noise sources (90dB, 80dB, 70dB, 60dB, 50dB)

In [1] we introduced an algorithm called “Acoustic eraser”. We were able to show, that a complex sound field can be decomposed to discrete components by an phase shifted subtraction of reconstructed source signals so that covered source become visible (Eq.2, Fig.2).

$$\boxed{y_i'(\vec{s}_1,t) = y_i(t+\Delta t(\vec{s}_1)) - \hat{f}(\vec{s}_1,t)} \tag{2}$$

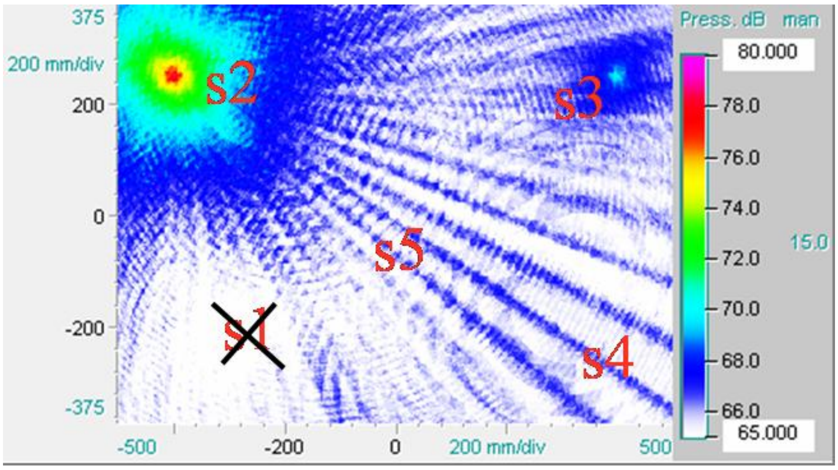

Fig.2: Acoustic photo with deleted source s1. Source s3 appears.

Based on (Eq. 3) an „inverse Acoustic eraser“ can be designed. The goal is to eliminate all sources except one (s1). The algorithm deletes all signals that are not correlated with the considered source. That leads to a separation of one source (s1) while all other source will disappear from the acoustic photo (Fig. 3).

$$\boxed{y_i''(\vec{s}_1,t) = y_i(t) - y'(\vec{s}_1,t)} \tag{3}$$

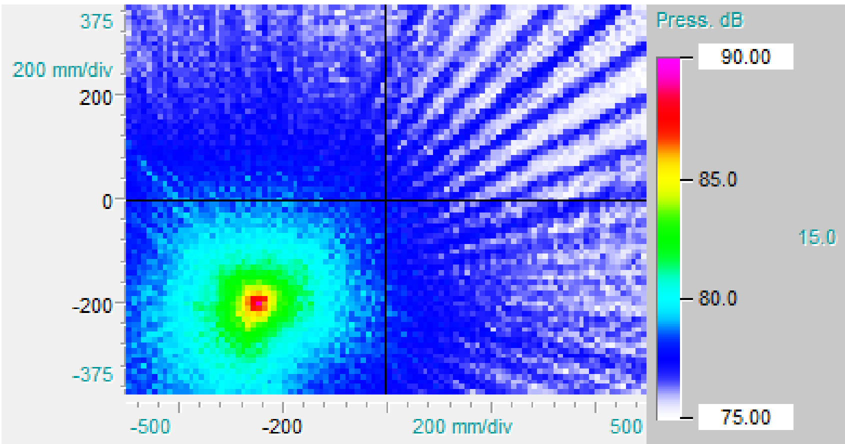

Fig.3: Acoustic photo with extracted source s1. s2, s3, s4, s5 are removed

The underlying problem is similar to a known problem of the classic photography: In a scene with high contrasts the optical system cannot completely reproduce the differences in brightness. The dynamic range of an optical system is greater than in the acoustics and reaches ratios from $1{:}10^{5}$ up to $1{:}10^{10}$ (80 dB to 180 dB) [5], [6]. Due to the choice of aperture areas of the photography are underexposed or overexposed and come along with a loss of details. **H**igh **D**ynamic **R**ange imaging describes a set of techniques to get a higher dynamic range between the darkest and lightest areas of a picture. Taking multiple photos of a scene with different exposure levels and intelligently merging these photos will lead to a picture that is representative in both dark and bright areas. A similar technique can be applied to acoustic photos by using the acoustic eraser. Following the HDR technique in the classic photography the algorithm is called HDR as well. In contrast to the high dynamic range imaging only one recording is needed to create high dynamic range acoustic photos, because the dynamic range of the microphones is sufficient. The HDR algorithm splits the microphone signals automatically into partial sources, which will be mapped separately and combined to one acoustic photo.

2 HIGH DYNAMIC RANGE ALGORITHM

The acoustic HDR technique is an iterative algorithm that works similar to CLEAN [7]. But in contrast to CLEAN it works in time domain, so known problems with windowing and average determination in frequency domain will be omitted. Hence, this technique is predestined for broadband signals, which cannot be separated in frequency domain. A separation of the results in frequency bands can be done by applying band pass filters to the microphone signals.

In every cycle the signals of the most dominant source including its sidelobes will be extracted. The sidelobes will get eliminated and the source will be added to the result photo. Following this step a new acoustic photo will be calculated to determine the next dominant source. One stop criterion for this algorithm is the maximum count of sources to be extracted.

1. $\hat{p}_{resulteff}(\vec{x}_j) = 0$

2. $\hat{f}(\vec{x}_j, t) = \frac{1}{M}\sum_{i=1}^{M} w_i y_i \left(t - \Delta t(\vec{x}_j)\right)$

3. $\hat{p}_{eff}(\vec{x}_j) = \sqrt{\frac{1}{n}\sum_{k=0}^{n-1} \hat{f}^2\left(\vec{x}_j, t_k\right)}$

4. $y_i'(\vec{x}_{\max}, t) = y_i(t + \Delta t(\vec{x}_{\max})) - \hat{f}(\vec{x}_{\max}, t)$

5. $y_i''(\vec{x}_{\max}, t) = y_i(t) - y'(\vec{x}_{\max}, t)$

6. $\hat{f}_{pm}(\vec{x}_j, t) = \frac{1}{M}\sum_{i=1}^{M} w_i y_i'' \left(t - \Delta t(\vec{x}_j)\right)$

7. $\hat{p}_{pm_{eff}}(\vec{x}_j) = \sqrt{\frac{1}{n}\sum_{k=0}^{n-1} \hat{f}^2\left(\vec{x}_j, t_k\right)}$

8. $N = \frac{1}{\hat{p}_{p\,eff}(\vec{x}_{\max})}$

9. $\hat{p}'_{pm_{eff}}(\vec{x}_j) = \hat{p}_{pm_{eff}}(\vec{x}_j) \bullet N$

10. $\hat{p}''_{pm_{eff}}(\vec{x}_j) = \left(\hat{p}'_{pm_{eff}}(\vec{x}_j)\right)^{sse}$

11. $\hat{p}'''_{pm_{eff}}(\vec{x}_j) = \frac{\hat{p}''_{pm_{eff}}(\vec{x}_j)}{N}$

12. $\hat{p}_{resulteff}(\vec{x}_j) = \hat{p}_{resulteff}(\vec{x}_j) + \hat{p}'''_{pmeff}(\vec{x}_j)$

13. $y_i(t) = y_i'(t)$

14 Next iteration starts at point 2.

3 EXAMPLES

The following images (Fig. 4) illustrate acoustic photos of a simulated acoustic scene. The image on the right was calculated using the HDR algorithm. The left one shows a standard beamforming result. The colored scale is applied to both images.

3.1 Simulations

All simulations were performed using a 32 channel microphone array with a ring geometry and a sampling rate of 48 kHz per channel. All sources were located at a distance of 1 meter from the microphone array.

3.1.1 Dynamic of HDR-algorithm

The first simulation was performed to separate the 5 incoherent sources (white noise) from figure 1 with levels of 90dB, 80dB, 70dB, 60dB and 50dB. Using the standard beamforming algorithm only the two dominant sources (s1, s2) can be determined. Using the HDR algorithm all source can be separated with their correct levels (max. failure 2dB) and location (difference only a couple of mm) (Fig.4). In contrast to the standard beamforming and under optimal conditions (anechoic room, broadband signal sources) the dynamic range is improvable by a factor of 300 (in this simulation from 10dB to 60dB). By increasing the resolution (quantization) of the microphone signals (in this simulation 16 Bit) more precisely results are conceivable.

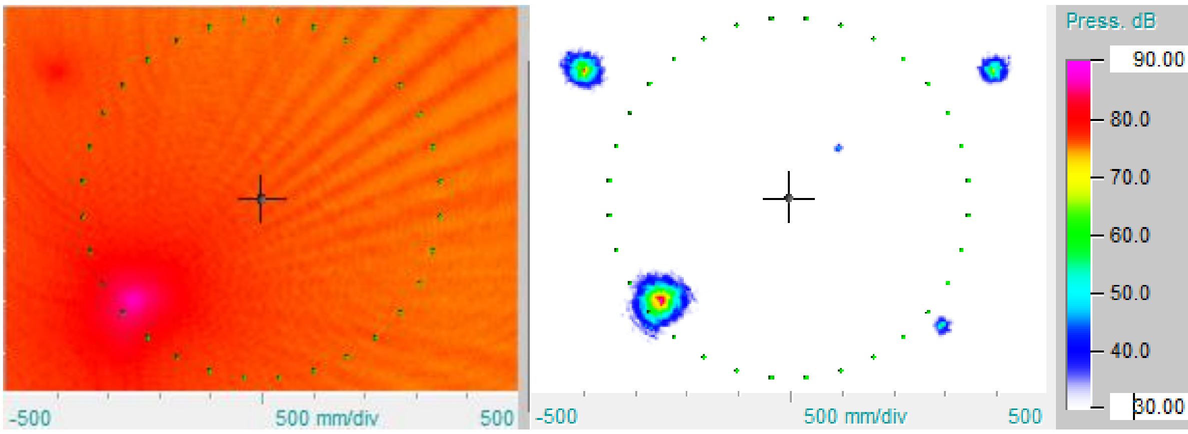

Fig.4: Acoustic photo of 5 incoherent noise sources without (left) and with HDR-algorithm

3.1.2 Coherent sources

Some source separation techniques (orthogonal Beamforming, [8]) provide acceptable results for independent sources, but they refuse to work in case of determining coherent sources (same source mechanism). In the following example (Fig. 5) the 5 sources are temporally coherent. Figure 5 illustrates that the HDR algorithm works reliable even with coherent sources.

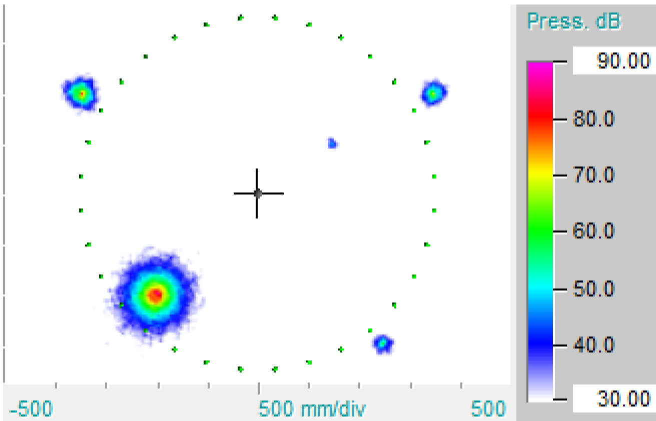

Fig.5: Acoustic photo of 5 coherent noise sources with HDR-algorithm (Dynamic: 60 dB)

3.1.3 Source separation

The HDR also provides a more accurate separation of adjacent sources. The next example shows that characteristic by simulating two sources (white noise, 90dB and 82dB). The sources are at a distance of 0.4m so they can be separated with help of the standard beamforming (Fig. 6). When the sources moving closer to a distance of 0.05m the sidelobes of the dominant source covers the weaker source completely. The HDR succeeded in separating both sources (Fig. 7).

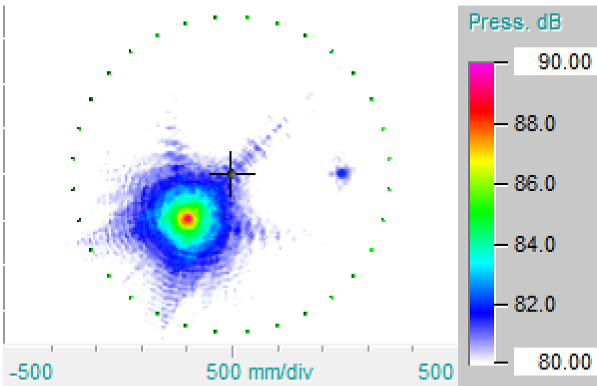

Fig.6: Acoustic photo of 2 noise sources (90 dB and 82 dB, distance between sources: 0.4m) with Standard-Beamforming

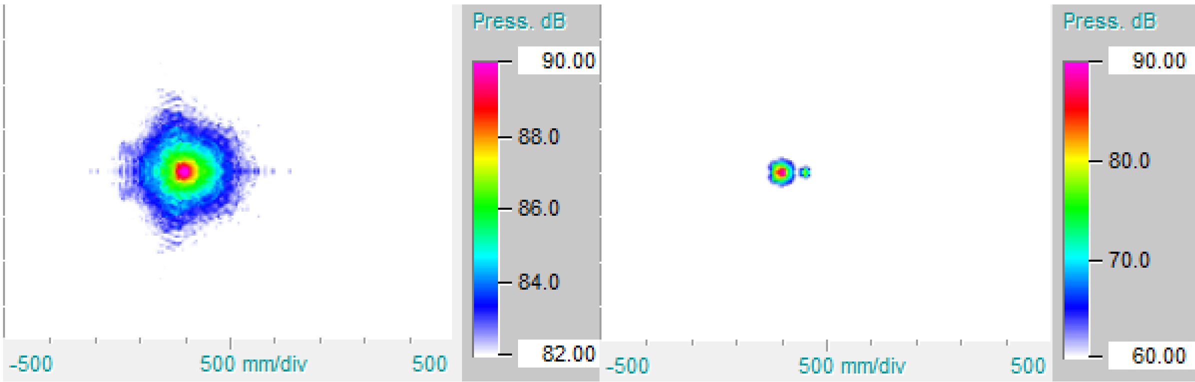

Fig.7: Acoustic photo of 2 noise sources (90dB, 82dB, distance between sources: 0.05m) with Standard-Beamforming (left) and HDR (right)

3.1.4 Signal reconstruction of quit sources

The HDR method allows a separation of low sources as well as a very accurate signal reconstruction with a slight portion of crosstalk of nearby dominant sources. The simulation shown in (Fig. 8) was performed with a sine-signal (2 kHz, 50dB) which is fully covered by 4 adjacent powerful sources (90dB, 80dB, 70dB, 60dB). With the help of the HDR algorithm this sine source can be reconstructed.

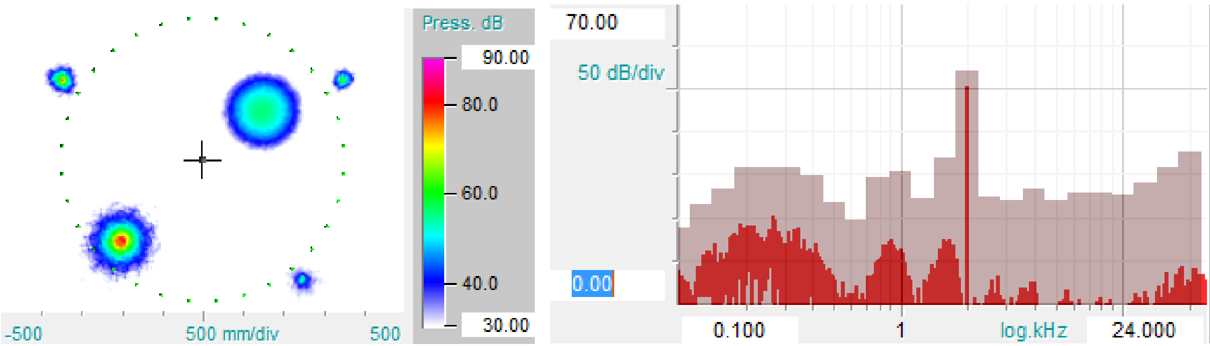

Fig.8: Signal reconstruction with HDR of a 2 kHz source (50 dB, middle) super positioned by 4 adjacent powerful noise sources (90dB, 80dB, 70dB and 60 dB)

With further examples we can demonstrate, that music and voice signals with levels up to 50dB below the ambient noise level can be reconstructed in a high quality.

3.2 Examples from practice

All measurements were performed with a 48 channel array with a ring geometry and different sampling rates (48 kHz – 96 kHz).

3.2.1 Anechoic room

The most accurate HDR results can be achieved in anechoic rooms. Figure 9 illustrates that using the example of an engine in a test stand. The single sources are clearly resolved with an increasing dynamic range from 10dB (conventional beamforming) up to 30dB.

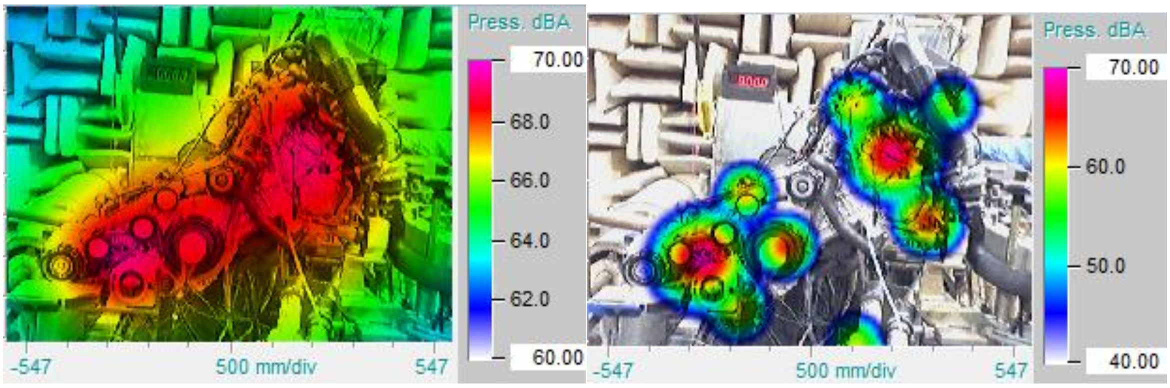

Fig.9: Acoustic photos of an engine under test. Left picture standard beamforming (10 dB dynamic), right side HDR (30 dB dynamic),

In a reverberant environment the HDR algorithm identifies all mirrored sources as well as sources from adjacent objects as real sources. Once they are located outside the image or measuring plane they cannot be focused correctly so they cannot be deleted from the microphone signals. Neverthless there is a clear enhancement in image quality compared to the results of conventional beamforming (Fig. 10).

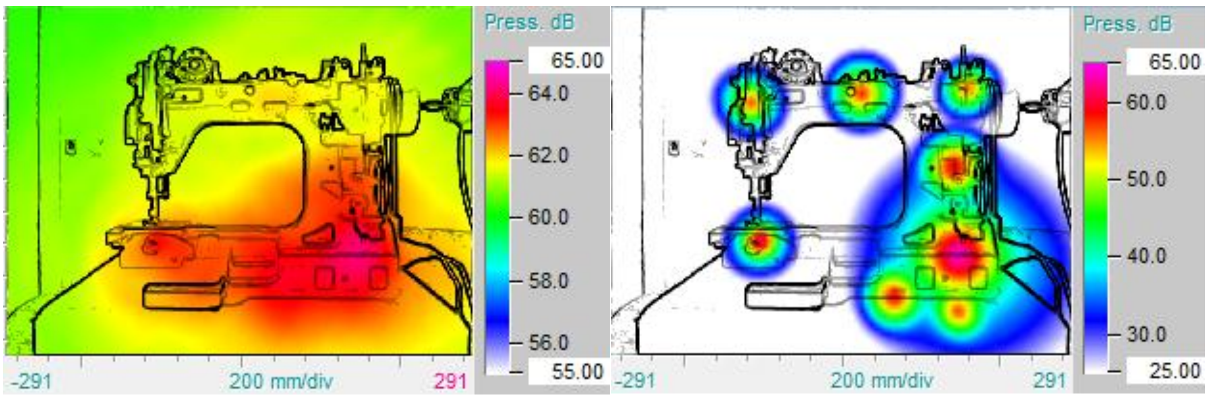

Fig.10: Sewing machine under test. Left picture standard beamforming (10 dB dynamic), right side HDR (40 dB dynamic)

4 CONCLUSIONS

The High Dynamic Range method extends the standard beamforming. This paper pointed out that the real contrast in acoustic scenes can be increased up to 60 dB under optimal conditions (anechoic room, broadband signals). Signals which are 50 dB below dominant adjacent sources can be reconstructed in a high quality. Coherent sources as well as incoherent sources can be seperated clearly. Compared to the standard beamforming the computing time for the HDR algorithm increases 10 to 20 times depending on the number of iterations.

$\hat{f}$	Approx. time function
$\vec{x}$	Map point
j	Map point index
t	time
M	Number of microphones
w	Weighting factor
y	Microphone signal
i	Microphone index
n	Number of samples
k	Sample index
$\max$	Indices of point with max. sound pressure in the map
pm	Part map
$\hat{p}_{pmeff}$	RMS of part map
N	Normalize factor
SSE	Sidelobe suppression exponent

5 ACKNOWLEDGEMENT

This work has been funded by the German Federal Ministry of Economics and Technology (BMWi) under the project registration number MF100061.

REFERENCES

[1] D. Döbler, Dr. Ralf Schröder: Contrast improvement of acoustic maps by successive deletion of the main sources, BeBeC 2010

[2] D.H.Johnson, D.E Dudgeon : Array Signal Processing. Concepts and Techniques, PTR Prentice Hall, 1993

[3] Sandro Guidati, Verfahren und Vorrichtung zur bildgebenden Darstellung von akustischen Objekten, Offenlegungsschrift DE102008026657A1 Deutsches Patent- und Markenamt

[4] Marcus Kern, Ein Beitrag zur Erweiterung von Beamfoming-Methoden, Dissertation TU Berlin 2008

[5] Reinhard u. a.: *High Dynamic Range Imaging,* S. 187

[6] J. A Ferwada: *Elements of Early Vision for Computer Graphics.* IEEE Computer Graphics and Applications 21, 5 (2001): 22–33, ISSN 0272-1716.

[7] Sijtsma, P., CLEAN based on spatial source coherence, International Journal of Aeroacoustics, Vol. 6, No. 4, pp. 357-374, 2007.

[8] Ennes Sarradj, Christian Schulze, Andreas Zeibig: *Einsatz eines Mikrofonarrays zur Trennung von Quellmechanismen*