Acquisition of exterior multiple sound sources for train auralization based on beamforming

Fanyu Meng, Frank Wefers, Michael Vorlaender
Institute of Technical Acoustics, RWTH Aachen University, Aachen, Germany

Summary
The auralization of train exterior noise requires not only the identification of different sound sources, but also the separated signal from each source. A virtual sound could not sound convincing if it is created by pure synthesis method without actual recordings, especially when there are multiple moving sources. Beamforming has been proved qualified to solve the identification and localization problems of multiple moving sound sources from trains, however, for auralization these information are not enough. This paper focuses on capturing the sound source signals from a running train by the method of beamforming for auralization. Beamforming is applied to identify the major sound sources and obtain their spatial distribution, and the recordings are used as actual signals for the synthesis as well. A microphone array is mounted along the railway side to carry out the measurement, which includes measuring the exterior noise from different types of trains at different speeds. With subsequent post-processing, the main sound sources are localized and identified. According to the spectra of the recordings, the sound sources are allocated separately to tonal and broadband components. The tones are generated by the propulsion system, while wheels, pantograph and coaches contribute mainly to the broadband part, namely rolling noise and aerodynamic noise. Finally, it is discussed how these individual signals can serve for the auralization of the complete train.

PACS no. 43.58.+z, 43.60.+d

1. Introduction

The noise generated by the high-speed vehicles not only has been the main annoyance to people at home but also to the outside environment. Auralization is understood as the process of turning predictive or measured acoustics data, such as a noise spectrum, into an audible audio signal in Virtual Reality (VR) [1]. With the synthesis of the sound sources or recording of trains, a virtual scenario with a train running will be realized, allowing the simulation and plausible experience of the generated noise. Therefore, the acquisition of the sound sources is the prerequisite for auralization.

Beamforming is an efficient and popular method to localize sound source [2] [3], including moving sound sources, such as aircrafts [4], trains [5], cars [6] etc. Acoustic beamforming uses microphones to form an array to enhance the detected audio signal. The geometry and weighting of the array determine certain performance, resolution, signal-to-noise (SNR) ratio, directivity etc. When the received signal is spatially filtered, the signal from a particular steered angle is enhanced, which is usually used for the unknown source localization. Based on the localization of the sound sources and their characteristics (frequency components, amplitudes, phases etc.), can the synthesis be conducted. For aircraft noise synthesis, a complete process is developed by Stephen A. Rizzi et al. [7], while for electrical railbound vehicles, a sound synthesis and validation model is proposed by M. Klemenz in [8].

This work uses a beamforming method to identify the sound sources on a train during passing by, which is the first step in the whole auralization procedure. Furthermore, the noise components are analyzed and the steered audio files are generated using beamforming as well. Finally, the following synthesis of pass-by noise for the train auralization in virtual environment is discussed.

2. Beamforming method

2.1. Delay and sum beamforming theory

When an array of microphones is used simultaneously to record a signal, the signals measured by each microphone can be denoted as $f(t, \mathbf{p}_n)$, where $\mathbf{p}_n$ is the coordinates of nth microphone. If we apply a time delay τ_n to a microphone compared to the origin of the

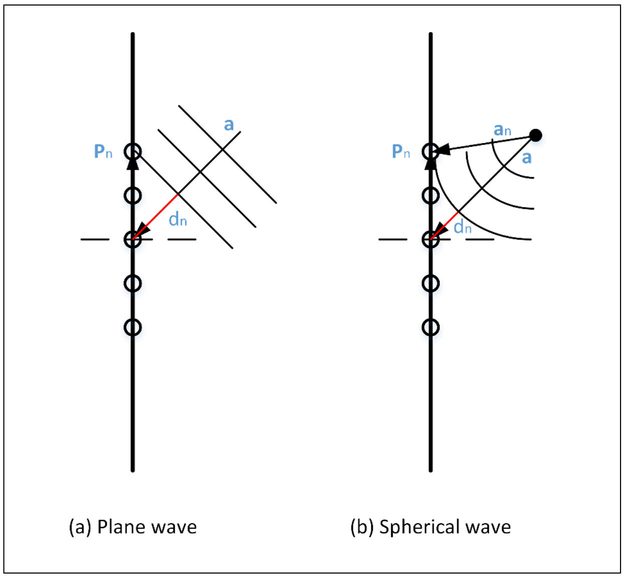

Figure 1. Traveling distance difference for plane and spherical waves.

array, the signal of this microphone is the same as that measured by the origin. The origin is assumed at the gravity center of the array and whether there is a microphone at the origin does not make difference. This time delay differs between plane and spherical waves because of different traveling distance difference d_n compared to the origin for the same microphone. Figure 1 shows the difference between the two types of waves.

τ_n is expressed as

$$\tau_n = \frac{d_n}{c} \tag{1}$$

where c is the speed of sound. For plane wave, $d_n = \mathbf{a}^T \mathbf{p}_n$, where $\mathbf{a}$ is the unit vector of the incident wave direction. Therefore,

$$\tau_n = \frac{\mathbf{a}^T \cdot \mathbf{p}_n}{c} \tag{2}$$

For spherical wave, $d_n = \|\mathbf{a} + \mathbf{p}_n\| - \|\mathbf{a}\| = \|\mathbf{a}_n\| - \|\mathbf{a}\|$, so

$$\tau_n = \frac{\|\mathbf{a}_n\| - \|\mathbf{a}\|}{c} \tag{3}$$

With all signals received by the microphones delayed with the corresponding delay time τ_n, the output of the array is achieved by summing them up. The measured signal is hence enhanced in the output compared to measuring only with one microphone. This algorithm is called delay and sum beamforming, or conventional beamforming [9].

$$y(t) = \sum_{n=0}^{N-1} w_n f(t - \tau_n, \mathbf{p}_n) \tag{4}$$

w_n represents the weighting multiplied on the output signal from the nth microphone and N is the microphone number. If w_n is not uniform, then the array is "shaded" [5]. The weighting is very important in adjusting the beamwidth and sidelobe level for a given array geometry, which will be discussed later.

If the output is transformed into frequency domain by Fourier Transform, time delay τ_n is then expressed by linear phase shift as $\phi_n = e^{-j\omega\tau_n}$ [10]. For plane wave,

$$\phi_n = e^{-j\mathbf{k}^T \mathbf{p}_n} \tag{5}$$

where $\mathbf{k} = \frac{\omega}{c}\mathbf{a}$ is the wave number. For spherical wave

$$\phi_n = \frac{e^{-jk\|\mathbf{a}_n\|} e^{jk\|\mathbf{a}\|}}{\|\mathbf{a}_n\|} \tag{6}$$

where $k = \frac{\omega}{c}$ is the magnitude of wave number and $e^{jk\|\mathbf{a}\|}$ is a constant. Besides, the phase shift is normalized with $\|\mathbf{a}_n\|$ as what has been done in the plane wave phase shift equation.

With a target direction $\mathbf{k}_T(\mathbf{a}_T)$, defining

$$\mathbf{v}(\mathbf{k}) = [e^{-j\mathbf{k}_T \mathbf{p}_0}; e^{-j\mathbf{k}_T \mathbf{p}_1}; ...; e^{-j\mathbf{k}_T \mathbf{p}_{N-1}}] \tag{7}$$

for plane wave, and for spherical wave

$$\begin{aligned}\mathbf{v}(\mathbf{k}) = [&\frac{e^{-jk\|\mathbf{a}_0\|} e^{jk\|\mathbf{a}_T\|}}{\|\mathbf{a}_0\|}; \frac{e^{-jk\|\mathbf{a}_1\|} e^{jk\|\mathbf{a}_T\|}}{\|\mathbf{a}_1\|}; \\ &...; \frac{e^{-jk\|\mathbf{a}_{N-1}\|} e^{jk\|\mathbf{a}_T\|}}{\|\mathbf{a}_{N-1}\|}]\end{aligned} \tag{8}$$

as the array manifold vector, the output of the array in frequency domain can be written as

$$Y(\omega, \mathbf{k}) = \mathbf{w}^T(\mathbf{k})\mathbf{F}(\omega) \tag{9}$$

where $\mathbf{w}_n(\mathbf{k}) = w_n \mathbf{v}_n(\mathbf{k})$, and $\mathbf{F}_n(\omega)$ is the Fourier Transform of $f(t, \mathbf{p}_n)$. This function is also defined as the frequency-wavenumber response function. If we specify the wave traveling direction, the array beam pattern can be obtained according to this function

$$B(\omega : \theta, \phi) = Y(\omega, \mathbf{k}) \mid_{k=\frac{\omega}{c}\mathbf{a}(\theta,\phi)} \tag{10}$$

Beam pattern plays a very important role in determining array performance, which will be discussed in the following section.

Figure 2. On-site measurement setup.

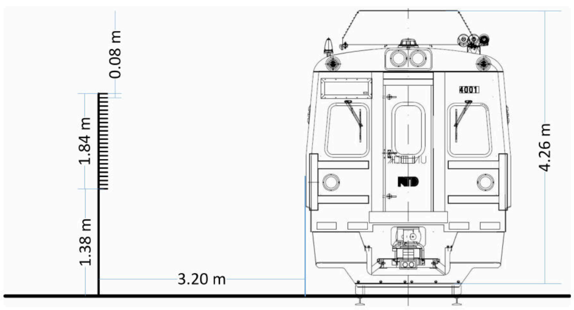

Figure 3. Front view of the on-site measurement setup.

3. On-site measurement and array performance

3.1. Measurement setup

A vertical linear microphone array with 24 uniformly spaced (0.08 m) microphones is positioned 3.2m from the near-side surface of the train, which can be seen in Figure 2 and Figure 3.

Table I shows the detailed setup information. Two types of regional trains, RE9 and RB20, in North Rhine-Westphalia, Germany are measured at speed of 150 km/h and 91 km/h. Except for the microphones, a video camera is also used to take videos during the train passing by. With subsequent synchronization of video and audio from the microphones, the relative position between the train and array can be seen in the video for a specific time, which helps localize find source position.

Table I. Microphone array setup parameters (D: array diameter, d: spacing between two microphones, L: distance to the train near-side surface, H: array center height)

D	d	L	H
1.84 m	0.08 m	3.20 m	2.64 m

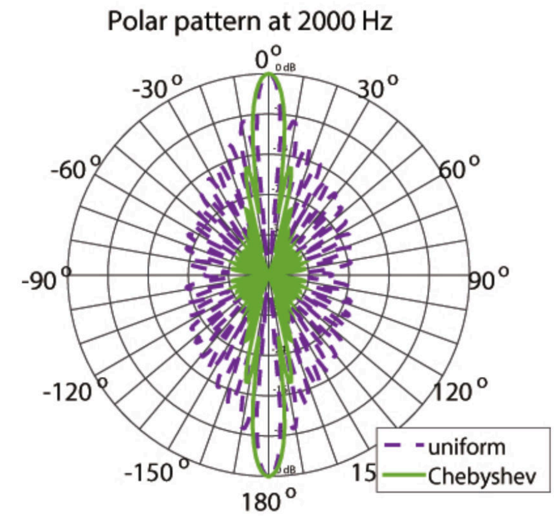

Figure 4. Beam pattern of uniform and Chebyshev weighting array (steering angle equals zero degree).

3.2. Array performance

The array beam pattern in terms of the given setup at 2 kHz with two different weighting is shown in Figure 4. From the beam pattern, important array performance parameters, maximum sidelobe level (MSL), half-power beamwidth ($HPBW$), beamwidth null-to-null (BW_{NN}) etc. can be acquired. It can be seen that using Chebyshev weighting increases the beamwidth but lowers the maximum side lobe level. This is always the case when a nonuniform weighting is applied.

The Rayleigh resolution, which is related to BW_{NN} can be calculated by equation [11]:

$$R(\theta) = 1.22\frac{L}{D}\lambda\frac{1}{cos^3(\theta)} \tag{11}$$

where L is the distance from the array center to near-side surface of the train, D is the array diameter, θ is the steering angle from the main response axis (MRA) of the array and λ is the wavelength. The resolution indicates that if two sound sources are within the range, they cannot be distinguished with each other. For instance, the resolution of this array at 2 kHz, 3.2 m distance, with no steering is 0.36 m (6°, steering angle equals zero).

As stated before, the weighting w_n is important to array performance. If uniform weighting is used, the resolution can reach the best when the noise at each microphone is assumed as spatially uncorrelated. Therefore the resolution and the sidelobe level reduction are reciprocal with each other. If the Chebyshev weighting is applied, the MSL increases by 9 dB to 18 dB and at the same time half BW_{NN} increases from 3° to 9° (0.5 m, steering angle equals zero) at 2 kHz compared to the uniform weighting. Taking Chebyshev weighting as an example of the nonuniform weightings is because it increases the beamwidth with the least value and the MSL with only around 1 dB less than the others.

4. Calculation and results

4.1. Beamforming on moving sound sources

The microphone array is vertical line array, so it does not have any resolution capability in the horizontal plane. However, spherical wave is taken as the propagating wave considering the distance from the train to the array. Therefore for a sound source on the train, the pressure received by microphones reaches the highest amplitude when the source position is right in front of the array. Therefore, the audio can be divided into many blocks with equal time slot, and for each time slot there is a corresponding part of the train "facing" the array. "Facing" means the center of this part is just in front of the array, or rather, in the resolvable area. Of course in a specific time slot the audio consists of the sound signals radiated from all the sound sources. However, a sound source could still be identified comparing with the sound pressure level in the neighbor blocks when it is "facing" the array.

Imagine there is a plane which sticks to the train near-side surface and the sound sources are all on this plane, on which the beamformer output is calculated and the sources are localized. This plane is called reconstruction plane [6], which is 3.2 m away from the array. The array focus can be steered from top to bottom for the part "facing" the array on the reconstruction plane. This part can be divided into grids, each grid representing a steering angle. As the train moves, the following parts are divided under the same rule, making the whole reconstruction plane into a mesh plane. Thereby a color map is generated where the grid with greater amplitude among the neighbor grids may indicate a sound source.

A simulation is run using the Matlab ITA-Toolbox developed at the Institute of Technical Acoustics at RWTH Aachen University [12]. Three moving sources are located at three different positions: a (1,3,0), b (0,3,5), c (-2,3,7) on a plane, and it moves at speed of 1 to the negative z direction. The microphone array is placed on the x axis from -0.92 to 0.92 with 0.08 spacing as in the real measurement (Figure 5). The

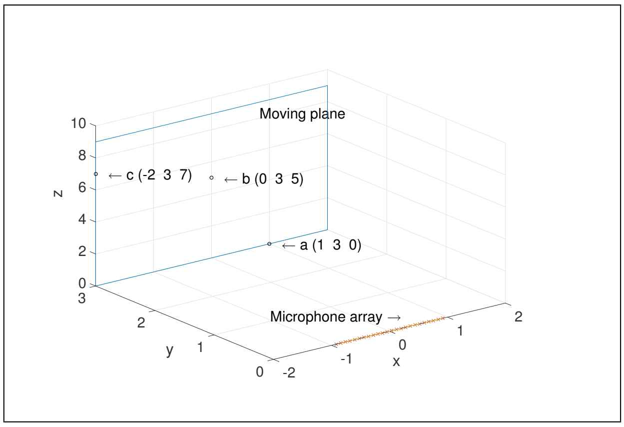

Figure 5. Microphone array and original positions of the moving sources in the simulation

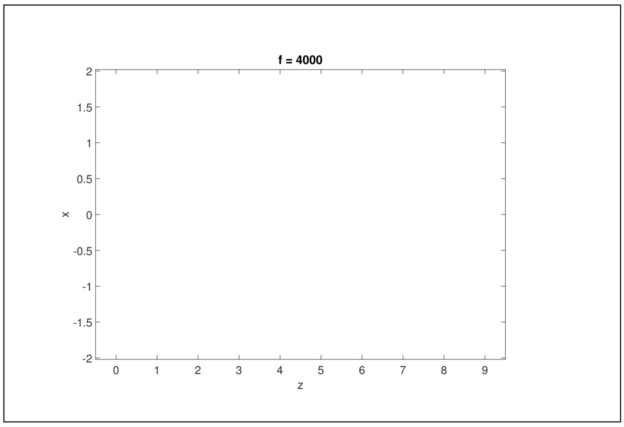

Figure 6. Simulation: color map of three moving sound sources.

calculated color map of the reconstruction plane at 4 kHz using the method above is shown in Figure 6. The sources can be clearly identified according to the colors. The areas also with stark colors near the sources illustrate that the array receives less energy from the particular source when it is not right in the front, and influenced by other sources as well. Even in each time slot the signals from all sources mixed up, they can still be isolated.

4.2. Sound source localization

The reconstruction plane for RE9 is partitioned into 529 × 460 grids, where the 529 is along the length of the train with 0.48 m spacing, corresponding to 512 samples of each audio block ; 460 in the elevation and 0.015 m spaced, which guarantees no spatial aliasing up to 5 kHz.

Uniform weighting is applied in the frequency below 2.5 kHz while for higher frequencies Chebyshev weighting is used, which can increase the MSL for higher frequencies. This is reasonable because using the Chebyshev weighting in lower frequency will lead to BW_{NN} improvement hence reducing the resolution. At 2.5 kHz, the resolution using Chebyshev

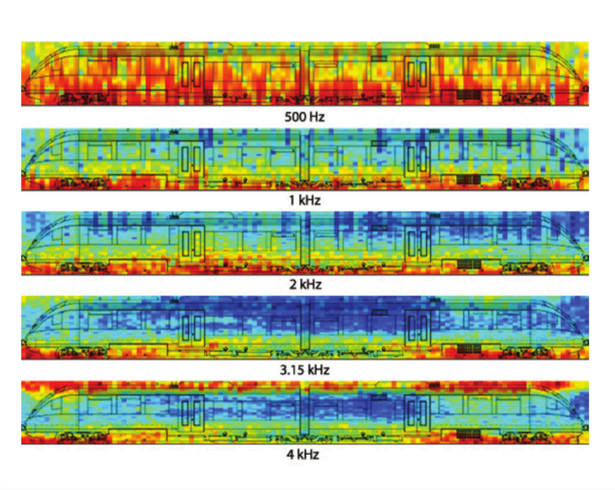

Figure 8. Source map on RB20 reconstruction plane.

weighting is similar to that using the uniform weighting at 2 kHz which is calculated previously, and the resolution increases with frequency. Thereby, the resolution is high enough using Chebyshev weighting above 2.5 kHz.

Figure 7 presents the source maps of RE9 at different 1/3 octave bands (the train is running to the left). It is clear that the rolling noise generated by the wheel/rail contact is one of the main sound source. As the frequency increases, the aerodynamic noise gradually increases as well, mainly oriented from the pantograph, the gaps between coaches and also the facilities outside of the train above the roof. Anyway, the rolling noise is still the strongest-power broadband component.

Figure 8 shows the source maps of RB20 (train running to the left). Again rolling noise contributes the highest power, and the facilities outside the train above the roof generate noise as well. However, since there are not as many facilities as those in RE9, they contribute not too much. The difference for RB20 is that the engine and cooling fans radiate almost the same strong noise as the wheels at 4 kHz.

4.3. Passing-by noise components analysis

The spectrogram of RE9 pass-by noise is shown in Figure 9. Before and after the train passes, there are clear tones at around 1370 Hz and 1800 Hz, while the rest signal is just broadband noise. Basically the synthesis can be separated into two individual parts: tonal and broadband synthesis. The sources corresponding to the two tonal frequencies are unclear, yet the synthesis can still be proceeded by generating a sinusoidal signal; while the broadband synthesis can be generated by filtering white noise by 1/3 octave band [1].

4.4. Discussion about further application

The sound source localization and the spectrum analysis both serve for the audio synthesis and auralization. Only after knowing where to put the sources and the components can one auralize a realistic and convicing train pass-by audible scenario in a virtual environment.

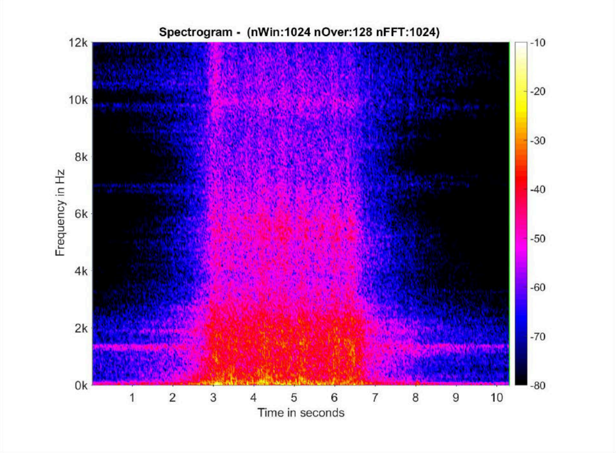

Figure 9. Spectrogram of pass-by RE9.

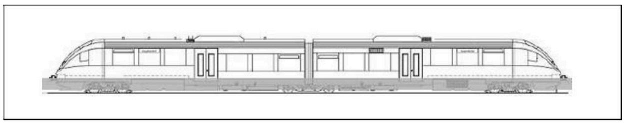

Figure 10. The "strip" on the train with a steering angle

Another application of the beamforming method is to get the steered audio files. Instead of allocating the audio into blocks, the whole pass-by audio is put in the beamforming algorithm and the output is the audio signals with different steering angles corresponding to a specific "strip" on the train. One example is shown in Figure 10 with a "strip" along the wheels. In this case, if the rolling noise is to be synthesized, the audio file steered to this angle gives frequency information more accurate than the original recording, which can be regarded as the reference signal. For other sources, choosing the corresponding audio file with the steered direction pointing to the source and the times lot when it passes the array will provide a good reference for the source synthesis. With the synthesized signals of sources and further study of their parameters, auralization can be conducted by controlling the sources and changing the parameters at real-time. In this way, people can evaluate what they hear during a train passing by with free movement.

5. CONCLUSIONS and OUTLOOK

The sound source localization for two pass-by trains is proposed using delay and sum beamforming and how to synthesize train pass-by noise is discussed as well. A method for source localization using vertical linear array based on conventional beamforming is developed, which works well through a simulation. The

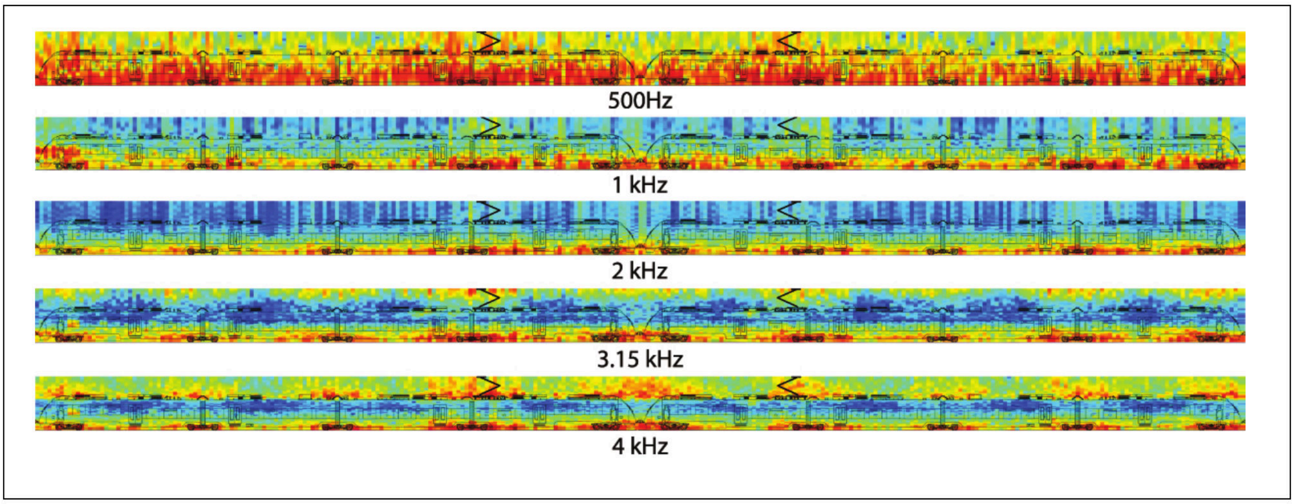

Figure 7. Source map on RE9 reconstruction plane.

main sound source of these two trains is the rolling noise generated by the wheel/rail contact. It is also pointed out that the aerodynamic noise from the pantograph and the coach gaps gradually increases with the frequency. Besides, engine and cooling fan are the main sound sources for RB20. The array performance in lower and higher frequency and the localization resolution will be improved when measuring with two-dimension and larger diameter array. Furthermore, pass-by noise synthesis can be conducted through tonal and broadband components individually and be applied in subsequent auralization. Further measurement is needed to increase the accuracy of the results, and extended application of beamforming will be discussed.

Acknowledgement

The authors acknowledge the Institute of Transport Science, RWTH Aachen University to support the information of the measured trains.

References

[1] A. Sahai, E. Anton, E. Stumpf, F. Wefers and M. Vorlaender. Interdisciplinary auralization of take-off and landing procedures for subjective assessment in virtual reality environments. 8th AIAA/CEAS aeroacoustics conference, 2012.

[2] H. Krim, M. Viberg. Two decades of array signal processing research: the parametric approach. Signal Processing Magazine, IEEE 13.4 (1996): 67-94.

[3] J. C. Chen, K. Yao, R. E. Hudson. Source localization and beamforming. Signal Processing Magazine, IEEE 19.2 (2002): 30-39.

[4] V. Fleury, J. Bultee. Extension of deconvolution algorithms for the mapping of moving acoustic sources. The Journal of the Acoustical Society of America 129.3 (2011): 1417-1428.

[5] B. Barsikow, W. F. King, E. Pfizenmaier. Wheel/rail noise generated by a high-speed train investigated with a line array of microphones. Journal of Sound and Vibration 118.1 (1987): 99-122.

[6] D. Yang, Z. Wang, B. Li, Y. Luo, X. Lian. Quantitative measurement of pass-by noise radiated by vehicles running at high speeds. Journal of Sound and Vibration 330.7 (2011): 1352-1364.

[7] S. A. Rizzi, B. M. Sullivan. Synthesis of virtual environments for aircraft community noise impact studies. 11th AIAA/CEAS Aeroacoustics Conference. Vol. 4. 2005.

[8] M. Klemenz. Sound synthesis of starting electric railbound vehicles and the influence of consonance on sound quality. Acta acustica united with acustica 91.4 (2005): 779-788.

[9] H. L. Van Trees. Optimum Array Processing: Part IV of Detection, Estimation and Modulation Theory. John Wiley Sons, 2002.

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

[11] D. Havelock, S. Kuwano, M. Vorlaender. Handbook of Signal Processing in Acoustics. Springer, 2008.

[12] M. Mueller-Trapet. Comparison of Sound-Source Localization Methods for Vibrating Structures. RWTH Aachen University, 2009.