DOI: 10.61782/fa.2025.0550

FORUM ACUSTICUM
EURONOISE 2025

MEASUREMENT-BASED SOUND SOURCE MODELLING OF MOVING VEHICLES FOR AURALIZATIONS UTILIZING ACOUSTIC BEAMFORMING

Thorben Kron[1*] **Daniel Johannes Meyer**[1] **Thomas Koch**[1] **Christoph Ende**[1]
Jonas Egeler[2] **Anton Schlesinger**[2] **Jens Bartnitzek**[3] **Laura Höhle**[3]
Benjamin Schlüter[4] **Ralf Böhme**[4]

[1] Fraunhofer Heinrich Hertz Institut, Berlin, Germany
[2] Möhler + Partner Ingenieure AG, Augsburg, Germany
[3] A+S Consult GmbH, Dresden, Germany
[4] Deutsche Eisenbahn Service AG, Putlitz, Germany

ABSTRACT

Auralizations of virtual scenes have become increasingly realistic in recent years, as respective sound propagation models and hardware platforms have made significant advances. 3D models of environments and digital twins have become more readily available due to the increasing adoption of digital planning methods, such as building information modeling (*BIM*). Here, auralizations offer great potential to experience the noise immissions of different scenarios. Despite their proven usefulness for public participation or psychoacoustic evaluation, there remains a deficiency in methodology to realistically simulate complex moving sound sources. For this reason, we present a new measurement-based approach to obtain highly realistic sound source models of moving vehicles. First, the vehicle is recorded with a microphone array to extract short audio signals from its dominant noise emissions through beamforming. To ensure robust source separation even at high velocities, we present an IP based microphone array that allows for high channel counts and flexible geometrical arrangements. The dominant emissions are then modeled as point sources. Their emission signals are calibrated, spectrally corrected, and extended in time using algorithms that ensure both physical and perceptive plausibility.

Keywords: *auralization, beamforming, microphone array, digital twin, EAV-Infra*

1. INTRODUCTION

In the planning of new railway lines, conflicts between local residents, operators, and other stakeholders regularly result in delays of several years. In particular, concerns regarding excessive noise exposure pose a significant threat to public acceptance. Modern simulation technologies make it possible to demonstrate future acoustic immissions realistically in advance, providing experts and laymen alike with an intuitive understanding of anticipated changes. Within our research project *EAV-Infra*, methods are being developed to convey acoustic changes in an intuitively comprehensible manner. In addition to the development of psycho-acoustic parameters for the subjective characterization of railway noise emissions [1] and the integration of acoustic calculations into digital planning environments, sound propagation models and sound source models are also being created for realistic spatial auralizations. This publication focuses on a novel method for measurement-based modeling of moving sound sources, which may consist of potentially numerous partial sources. A team of researchers developed

**Corresponding author*: *thorben.kron@hhi.fraunhofer.de.*

an alternative approach that follows a synthetic methodology whose main focus is on wheel-rail emissions [2]. However, this cannot be applied to other prominent sound sources that are dominant especially at higher velocities. The use of microphone array measurements for spatially accurate detection of railway noise has been a subject of research for several years [3, 4]. However, such investigations have typically emphasized the technical characterization of sound sources, whereas the reproduction of acoustic emissions in simulation environments has been considered to date primarily in theoretical terms [5]. We propose a new approach for the creation of digital acoustic models of complex moving sound sources. Chapter 2 outlines the methodology employed. First, a novel hardware concept for microphone arrays is introduced, enabling flexible, synchronous transmission of extremely high channel counts. Subsequently, the procedure for computing sound source models from recordings obtained with this sensor system is presented. Chapter 3 describes the practical application of the method using the example of a high-speed *ICE 4* train that was recorded on a German railway track.

2. METHODOLOGY

2.1 Microphone Array Development

2.1.1 Electrical and Mechanical Design

The microphone array system is composed of both electrical and mechanical designs that integrate network based transmission of audio signals, control data and power. The electrical and mechanical structure is centered around two printed circuit boards (*PCB*). The primary PCB is of octagonal design with a diameter of 45 cm and accommodates 64 *MEMS* microphones and essential peripherals. The microphones are arranged according to an optimized *Underbrink* spiral geometry [6] specifically tailored for railway noise analysis. Each microphone contains its own analog-to-digital converter. The smaller secondary control PCB manages power distribution, control signals, and network connectivity using standard *RJ45* connectors. The board interfaces with the main PCB through a pin header. Part of the main PCB design are eight dedicated integrated circuits (*ICs*) responsible for converting the pulse-density modulated signals from groups of eight microphones into pulse-code modulated data streams. The ICs receive their clock signals from a field-programmable gate array (*FPGA*) module located on the control PCB via buffered clock lines. The audio clock synchronization across the microphone array is centrally managed by the FPGA, which in turn synchronizes to the time server over the IP network. Audio data is likewise transmitted over IP protocols. Control data communication occurs through the I^2C protocol, while bidirectional serial data exchange for device configuration and control utilizes either *USB* or *TCP/IP* [7]. Power is delivered to the array through the use of *Power over Ethernet* (*PoE*). The entire PCB assembly can be fastened to tripods and other mechanical components via multiple mounting holes. In order to minimize wind effects and vibrations, each microphone array can be installed in a specially designed, doubly suspended windshield. Figure 1 shows a rendering of the primary PCB.

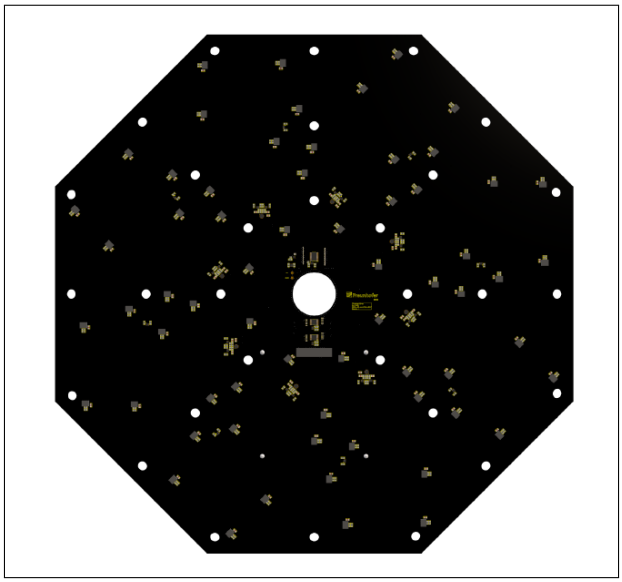

Figure 1. Rendering of the primary PCB.

2.1.2 Data Transmission

The system leverages audio streaming and synchronization via standard IP networks, ensuring interoperability, scalability, and efficient data management. The audio transmission relies on the *AES67* [8] standard, which facilitates Audio-over-IP interoperability by providing a method for transmitting digital audio across IP-based networks. AES67 utilizes the *User Datagram Protocol* (*UDP*) [9] as its transport mechanism, either in multicast, or unicast setups. To achieve robustness and mitigate potential packet losses inherent to UDP, the system integrates Quality of Service protocols. In practice, *Differ-*

entiated Services Code Points [10] are used to prioritize audio and synchronization packets within heterogeneous environments. The array system is synchronized to other AES67 servers and clients by the *Precision Time Protocol, Version 2* (*PTPv2*), compliant with the *IEEE 1588-2008* standard [11]. Revision 2023 of AES67 further requires *IEEE 1588-2019* compatibility. PTP ensures accurate and coherent clock alignment across networked devices by establishing a common timing reference via a hierarchical structure comprising different clock types. This approach enables time-aligned transmission and synchronized audio data reception, which is necessary for beamforming and other acoustic localization applications. Therefore, the synchronized and time-aligned parallel operation of multiple arrays in conjunction is made feasible. The *Real-Time Transport Protocol* (*RTP*) [12] is responsible for the audio data transport, embedding timestamp and sequence numbering information in the packets. The timestamps allow for stream synchronization at the receiver, while sequence numbers enable reordering of packets, as well as detection of packet loss. Session management within AES67 is performed using the *Session Description Protocol* [13], which facilitates the exchange and negotiation of audio stream parameters, such as codec configurations, IP addresses, sample rates, and synchronization settings. Although AES67 itself does not specify device discovery and advertisement methods, the *Session Announcement Protocol* [14] has been adopted by most vendors to simplify device management within an AES67-enabled network.

2.1.3 *Operation of Multiple Arrays in a Network*

As explained above, the operation of multiple arrays in parallel is an inherent feature of the system architecture, therefore only the definition of the multi-array geometry and its mechanical construction had to be considered. A multi-array setup extends the lower frequency limits of the array aperture. The geometry was influenced by the decision to limit the number of sub-arrays to 5, as experimental testing showed that the required spatial sampling bandwidth could be sufficiently realized. Further generative optimizations led to the geometry depicted in 2, where each sub-array is located on one of the vertices of a regular pentagon. Mechanically, each windshield is attached to an aluminum profile through 3D-printed mounting brackets. The profiles are then fastened to a steel plate, which in turn can be mounted on standard tripods with a *TV spigot*. The setup is depicted in figure 2.

Figure 2. Configuration of five arrays.

The array network can be treated like a single array, with each sub-array acting as an individual AES67 node. This results in a total of 320 microphones, as each sub-array comprises 64 sensors. The 320 audio signals are then transmitted to a computer through a central switch that connects all nodes. Although many AES67 devices can act as the so-called PTP *grandmaster*, we deploy a dedicated PTPv2 time leader within the system. This grandmaster acts as the common time reference for all nodes. In total, the system is therefore comprised of 5 arrays, a PoE-enabled switch, a PTP grandmaster and a computer. The latter receives the 320 signals as per AES67 through a *virtual soundcard*. The term describes a software solution that handles the AES67 interfacing on common network cards. The individual signals can then be recorded through audio frameworks specific to the host operating system, such as *Steinberg ASIO* on *Microsoft Windows*, or *ALSA* on *Linux*. Synchronization error between sub-arrays is a key parameter in a multi-array system. All sensors need to be clocked synchronously and errors manifest themselves in phase uncertainties, resulting in localization errors. When measuring clock offsets

between sub-arrays, we found the maximum error to be in the order of about 70 nanoseconds. Assuming a sinusoid signal, the resulting maximum radian phase error can be calculated as:

$$\text{Error}_\phi \equiv \Delta\phi_{\max} = 2\pi \cdot f \cdot t_{\text{offset}_{\max}} \tag{1}$$

With frequency f in Hertz and $t_{\text{offset}_{\max}}$ describing the maximum offset in seconds. Assuming an upper limit of hearing of 20 kHz, the maximum phase error thus translates to about $\pm$ 0.504 degrees:

$$\Delta\phi_{\max} = 2\pi \cdot 20\,\text{kHz} \cdot 7 \times 10^{-8}\,\text{s} \tag{2}$$
$$\approx 0.008796\,\text{rad} \tag{3}$$
$$\approx 0.504\,\text{deg} \tag{4}$$

2.2 Sound Source Modeling

2.2.1 Measurement Setup

In order to capture trains adequately for the presented method, we use the network of five microphone arrays described in 2.1.3 for sound source separation based on acoustic beamforming. Additionally, we use a free-field compensated measurement microphone to get a reference of the pass-by sound pressure levels and a dummy head for binaural reference signals, which serve for validation purposes. A camera placed in the center of the array network allows for the superposition of the acoustic map with actual images of the sound sources and can be used to calculate the source velocity. All sensors are placed equidistantly from the track.

2.2.2 Array Signal Processing

In the first step, the microphone array signals are calibrated in amplitude and the spectrally corrected, which is mainly necessary to reduce the influence of the pressure build-up effect at higher frequencies due to the rigid PCB surface. For the spectral correction, a sine sweep measurement is performed for each sub-array and a measurement microphone. The magnitude of each microphone is aligned to the magnitude of the measurement microphone which is assumed to have a sufficiently flat magnitude response. This is done with a linear phase third-octave filter bank so that the phase relations remain unchanged for the later beamforming processing.
In the next step, *delay and sum beamforming* is performed with a moving grid to account for the source velocity and to eliminate the Doppler shift. The beamforming algorithm is performed block-wise with an overlap of 50 % between neighboring windows. This procedure is similar to the one described in more detail in [5]. The delay and sum algorithm is calculated using *Acoular*, an optimized and established Python framework for acoustic beamforming developed at TU Berlin [15, 16]. Longer block-lengths result in longer output signals and therefore reveal more information about the sources, but extreme off-center angles induce uncertainties due to the horizontal source directivities and a reduced beamforming performance. The option of performing a CLEANT deconvolution [17] on the resulting acoustic map has been investigated to minimize the influence of side-lobes of the arrays' frequency dependent *point spread function* but has been found to negatively influence the quality of the resulting audio signals, although it offers great advantages for different use cases. The camera, which visually captures the moving sound source from the center of the microphone array, is synchronized with the audio signals. The single pictures are cropped horizontally, with a width according to the source motion per frame, and stitched together to obtain an image of the complete source which can be overlapped with the acoustic map. The beamforming results are stored in two ways: A broadband version that covers the entire audible frequency range is used for extracting the actual audio signals which are used as a basis for the partial sound sources. A band-limited version that contains only higher frequency components of around 4 kHz upwards is used to generate the acoustic map by fading the single frames of each block into each other. This map is used to analyze the SPL maxima and their geometrical positions to define adequate point source locations. While the visual, band-limited beamforming map is calculated with a fine grid size to get as many details as possible, the wide-band beamforming map has a rougher grid size as the real sources have a certain spatial extent beyond the idealized point source representation and their signals should be captured within single pixels of the grid at the best.

2.2.3 Signal Prolongation

Depending on the speed of movement of the sound source, the beamforming provides only short audio signals of the estimated partial sound source locations. However, for the simulation of the partial sound sources, an arbitrarily long emission signal is needed. Simple looping does not suffice, as repetitive sounds without variations are perceived as strongly artificial. Thus, several approaches were tested to generate an adequate emission signal. In a vocoder-based approach, a multiband energy envelope was derived from the short audio signal, which was then transferred

onto white noise as a carrier signal. However, white noise could not authentically capture the detailed timbre of the audio signal. In another approach, linear extrapolation was implemented as described by Kauppinen et al [18]. Convincing results are achieved if enough samples are provided, however it is computationally expensive. Another method commonly used in media production contexts such as sound design is called *granular synthesis*. Grains, which are audio excerpts of a random duration between 40 and 70 ms, are randomly selected from the beamforming result. The emission signal is then generated by randomly placing the layered grains in time. This approach also achieved convincing results. It captures the texture of the original sound, which is reflected in both the spectrum and the acoustic evaluation. However, periodic details and transients are often washed out by the layering of the grains. For trains as sound sources, this tradeoff seemed acceptable, as the noise is predominantly perceived as static. Its fast computation enables flexibility during testing. Thus, this algorithm was chosen for evaluation and processing purposes.

2.2.4 Source Model Correction Based on Simulations

With the prolonged beamforming results as emission signals the sound source models are completed by assigning basic directivities. By defining $A_f \in [0..1]$ for frequency f, the directivity s_f depending on the angle of incidence θ is defined as:

$$s_f(\theta) = A_f + B_f \cdot \cos\theta \tag{5}$$

$$\text{with} \quad B_f = 1 - A_f \tag{6}$$

This allows for frequency-depended directivity ranging from omnidirectional for $A_f = 1$ to a dipol for $A_f = 0$. For the wheel-rail-contact for mid to high frequencies a figure-8 pattern is chosen in coherence with the *Schall 03* [19]. Based on assumption the directivity for lower frequencies tends towards omnidirectionality. To evaluate the sound source models the original sound event is simulated and compared to the reference recordings. For the simulation, a basic sound propagation model is applied: Ground reflections are simulated as mirror sound sources with a frequency-dependent reflection factor in accordance with the dataset of the software CATT-Acoustic for 100 mm gravel floor. The Doppler effect is simulated with a delay line. The amplitude is corrected depending on the distance r with a factor of r^{-1}. Dissipation is taken into account according to the norm ISO 9613-1 [20]. For the binaural rendering a spherical far field HRIR from Bernschütz [21] was used. The sound source models are then adjusted by comparing the simulation to the reference. Most importantly, the spectra of the emission signals are corrected by a spectral comparison between the simulation and the reference.

3. PRACTICAL APPLICATION

3.1 Procedure

Figure 3. Measurement setup for capturing trains for sound source modeling.

The presented method was tested in practice by performing a measurement at the railway line 6107 in Germany just outside the city of Stendal. The site offered near free-field conditions without other prominent sound sources other than the trains (figure 3). The sensors were placed as close as possible to the track at a distance of 11.0 m. A high-speed train, the ICE 4, was captured with a velocity of 179.2 km h^{-1} and serves as the subject of our investigation. For the sound source modeling, we calculated the beamforming results with a blocksize of 12 m. For the geometrical analysis based on the acoustic map we used a grid size of 0.2 m and for the spatial filtering of the actual audio signals we used a grid size of 1 m which result in a length of 241 ms before the prolongation. The trains consists of eleven center coaches, one of which has a pantograph and two end cars. The resulting acoustic maps are depicted in figure 4. Every bogie of the train as well as the pantograph are clearly recognizable and modeled as point sources. However, the measurement configuration does not allow to clearly separate the individual

wheels in every pair. Aerodynamic sound sources are less suitable to be viewed as point sources but need to be considered as well at higher velocities. Sound pressure maxima occur at higher elevations mainly at the transitions between the single cars. As a matter of simplification they are also modeled as point sources at a height of 4.3 m.

3.2 Validation

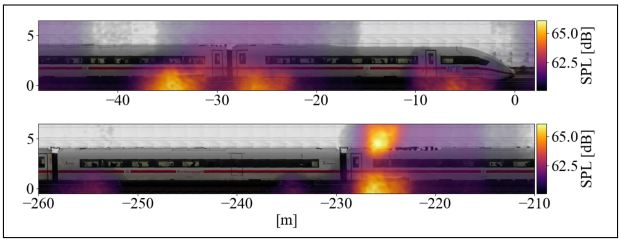

Figure 4. Sound pressure level map of captured ICE 4 train for frequencies from 4 kHz upwards (top: front of the train, bottom: pantograph emissions).

The simulation shows good agreement with the measurement regarding both analytical and subjective evaluation. With the spectral correction of the sound source models as described in section 2.2.4 the simulation's spectrum almost perfectly matches the original pass-by measurement, as figure 5 shows. Regarding the energy of the signals, figure 6 shows that both the maximum SPL and the mean SPL match closely. Over time, the energy peaks caused by the passing wheel-rail contacts seem to match and generally align as well. This suggests a general adequacy of the location and the number of the simulated sound sources. Compared to the measurement, the simulation shows a slightly stronger drop in energy between sound sources. This is probably caused by using strict point source models for only the main points of emission. A higher number of sound source models might achieve a higher precision, while possibly also introducing new challenges. The similar steepness of the rising and falling edge of the pass-by suggests a general appropriateness of the directivity of the sound source models. Informal listening tests by expert listeners confirm the results of the analytical evaluation. The loudness curve matches well, with regards to both the steepness of the edges at the beginning and end of the pass-by, and for the drops between the sound sources. The overall frequency distribution is perceived identical during the pass-by. Furthermore, the texture and coloration of the sound of the train are generally well captured. A slight coloring might be attributed to the granular synthesis prolongation procedure. Still some differences between the measurement and the simulation could be identified. The tonal component of the train noise seemed to be more prominent in the simulation, while the measurement seemed to include more noise components. This might be caused by aerodynamic noises, as they are highly prominent for high speeds, but can be modeled with point sources only to a limited extent. Before the pass-by there seems to be a soft rise in high frequencies, with a corresponding soft drop in frequencies after the pass-by. This could be caused by static rail noises, which are not included in the simulation. While already being very positively evaluated, the identification of differences by analytical and subjective evaluation is a valuable and effective means to further refine and adapt the simulation.

4. CONCLUSION

We presented a new measurement-based approach to auralize the pass by of a vehicle through extracting dominant emissions via beamforming and creating corresponding source sound models to acoustically simulate the original measurement. A network of proprietary IP based microphone arrays was used for recording. Using the aforementioned signal processing techniques, we extracted the sound source models and adapted the simulation, finally resulting in a perceptively convincing and physically plausible auralization of the original sound event. In future applications, geometrically optimized microphone distribution, as well as larger array networks, will further improve the performance of the source separation. Since the process was validated with the measurements it was based on, further research would be particularly enriching. The method used could be tested on new train configurations, different acoustic environments, and various sound sources to evaluate its general applicability.

5. ACKNOWLEDGMENTS

The authors acknowledge the support of the *Federal Ministry for Economic Affairs and Climate Action of Germany* for funding the project EAV-Infra. We also gratefully acknowledge *H7R* for providing their AES67 virtual soundcard and PTPv2 services *Streamer++* and *PTP++*, as well as hours of support. We further extend our thanks to *Cinela* for supplying the windshields and *Fachgebiet Technische Akustik* (*TU Berlin*) for providing Acoular and for their time and helpful feedback.

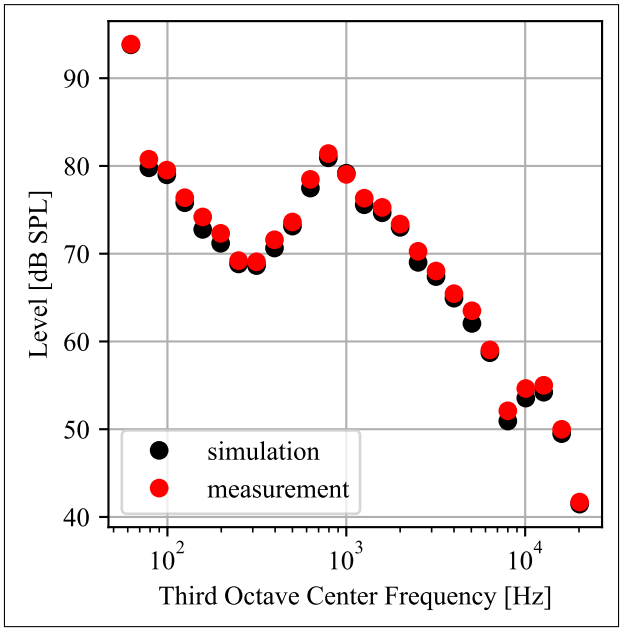

Figure 5. Comparison of spectra during pass-by.

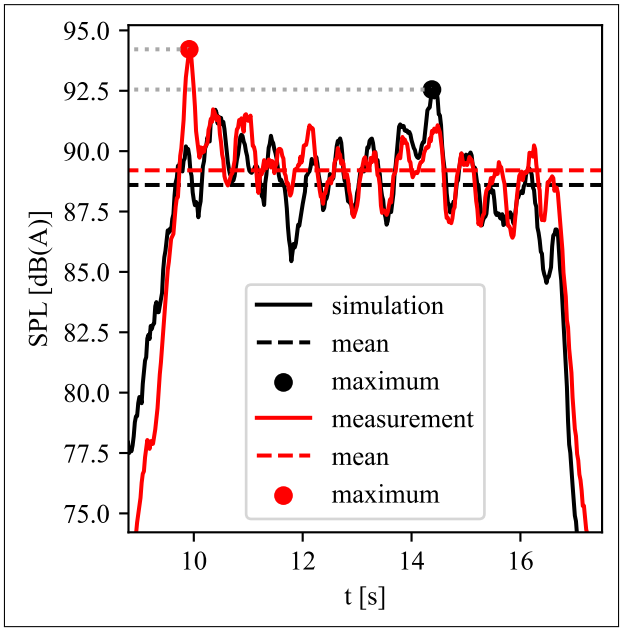

Figure 6. A-weighted SPL of the measurement.

6. REFERENCES

[1] Egeler, J., Huth, C., Schlesinger, A. et al., "Railway psychoacoustic annoyance - an indicator for planning infrastructure projects," *Proceedings of the Inter-Noise 2024 : 53rd International Congress and Exposition on Noise Control Engineering*, 2024.

[2] Pieren, R., Georgiou, F., Heutschi, K., Squicciarini, G., Thompson, D., "VR demonstration of railway noise mitigation using auralised train pass-bys," *Proceedings of the 10th Convention of the European Acoustics Association*, 2023.

[3] Kujawski, A., Sarradj, E., "Application of the CLEANT method for high speed train measurements," *Proceedings of the 8th Berlin Beamforming Conference 2020*, 2020.

[4] Zhang, J., Squicciarini, G., Thompson, D. J., Sun, W.,Zhang, X., "A hybrid time and frequency domain beamforming method for application to source localisation on high-speed trains," *Mechanical Systems and Signal Processing*, vol. 200, p. 110494, 2023.

[5] Meng, F., Masiero, B., Vorländer, M., "Compressive beamforming for moving sound source auralization," *Proceedings of the Inter-Noise 2016 : 45th International Congress and Exposition on Noise Control Engineering*, 2016.

[6] J. R. Underbrink, *Aeroacoustic Phased Array Testing in Low Speed Wind Tunnels*, pp. 98–217. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002.

[7] W. Eddy, "Transmission Control Protocol (TCP)." RFC 9293, Aug. 2022.

[8] Audio Engineering Society, AES67-2023, "AES standard for audio applications of networks - High-performance streaming audio-over-IP interoperability," 2023.

[9] J. Postel, "User Datagram Protocol." RFC 768 (Standard), Aug. 1980.

[10] F. Baker, D. L. Black, K. Nichols, and S. L. Blake, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers." RFC 2474, Dec. 1998.

[11] Institute of Electrical and Electronics Engineers, IEEE 1588-2008, "IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems," 2008.

[12] H. Schulzrinne, S. L. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications." RFC 3550, July 2003.

[13] A. C. Begen, P. Kyzivat, C. Perkins, and M. J. Handley, "SDP: Session Description Protocol." RFC 8866, Jan. 2021.

[14] M. J. Handley, C. Perkins, and E. Whelan, "Session Announcement Protocol." RFC 2974, Oct. 2000.

[15] Sarradj, E., Herold, G., "A Python framework for microphone array data processing," *Applied Acoustics, 116, 50–58*, 2016.

[16] Sarradj, E., Herold, G., Kujawski, A., Jekosch, S., Pelling, A. J. R., Czuchaj, M., Gensch, T., & Oertwig, S., "Acoular – Acoustic testing and source mapping software." Zenodo, https://zenodo.org/doi/10.5281/zenodo.3690794.

[17] Cousson, R., Lecl'ere, Q., Pallas, M.-A., Berengier, M., "A time domain clean approach for the identification of acoustic moving sources," *Journal of Sound and Vibration, 443, 47–62*, 2019.

[18] I. Kauppinen, J. Kauppinen, and P. Saarinen, "A method for long extrapolation of audio signals," *Journal of the Audio Engineering Society*, vol. 49, no. 12, pp. 1167–1180, 2001.

[19] Bundesamt für Justiz, "Sechzehnte Verordnung zur Durchführung des Bundes-Immissionsschutzgesetzes (Verkehrslärmschutzverordnung - 16. BImSchV), Anlage 2 (zu § 4), Berechnung des Beurteilungspegels für Schienenwege (Schall 03)," 2014.

[20] International Organization for Standardization, ISO 9613-1:1993-06, "Akustik; Dämpfung des Schalls bei der Ausbreitung im Freien; Teil 1: Berechnung der Schallabsorption durch die Luft," 1993.

[21] B. Bernschütz, "Spherical far-field HRIR compilation of the Neumann KU100," *Proceedings of the 39th Fortschritte der Akustik (DAGA)*, pp. 592–595, 2020.