MIRaGe: MULTICHANNEL DATABASE OF ROOM IMPULSE RESPONSES MEASURED ON HIGH-RESOLUTION CUBE-SHAPED GRID

Jaroslav Čmejla*, Tomáš Kounovský*, Sharon Gannot†, Zbyněk Koldovský* and Pinchas Tandeitnik†
* Acoustic Signal Analysis and Processing Group,
Faculty of Mechatronics, Informatics and Interdisciplinary Studies, Technical University of Liberec,
Studentská 2, 461 17 Liberec, Czech Republic.
Email: {Jaroslav.Cmejla, Tomas.Kounovsky, Zbynek.Koldovsky}@tul.cz
† The Alexander Kofkin Faculty of Engineering
Bar-Ilan University, Ramat-Gan, 5290002, Israel.
Email: {Sharon.Gannot, Pinchas.Tandeitnik}@biu.ac.il

***Abstract*—We introduce a database of multi-channel recordings performed in an acoustic lab with adjustable reverberation time. The recordings provide detailed information about room acoustics for positions of a source within a confined area. In particular, the main positions correspond to 4104 vertices of a cube-shaped dense grid within a $46 \times 36 \times 32$ cm volume. The database can serve for simulations of a real-world situations and as a tool for detailed analyses of beampatterns of spatial processing methods. It could be used also for training and testing of mathematical models of the acoustic field.**

***Index Terms*—Room Impulse Response, Acoustic Transfer Function, Microphone Array, Database**

I. Introduction

An exact mathematical description of the sound propagation in acoustic environments is difficult to define, as it is influenced by the shape of the room and by all objects, materials, and other physical properties of the enclosure. Since the propagation is linear, it is characterized by room impulse responses (RIR) or, equivalently, acoustic transfer functions (ATF) in the frequency domain. An RIR characterizes the sound propagation from one location to another.

When the RIRs relating some source positions and microphone array arrangements in a given room are known, the microphone signals can be generated with a high precision by filtering the emitted acoustic sources by the respective RIRs. This enables us to test signal processing algorithms and their behaviour in real-world conditions and, in particular, to verify the spatial robustness of multi-channel processors, to analyze acoustic fields and beampatterns, etc.

For simulating RIRs, it is popular to use generators based on the image method [1], which provide easy testing of algorithms as how they respond to changes of simple parameters such as reverberation time T_{60}, reflection coefficient, source distance or array geometry. However, generally the image method assumes artificial rectangular rooms ('shoebox') with simple walls containing no objects and no diffusive materials. The generated RIRs thus correspond to linear combinations of fractional-delay filters and fail to exactly imitate RIRs of real rooms with similar shapes.

Realistic RIRs can be obtained from RIR databases that were measured in real conditions;[1] see e.g., [2], [3], [4], [5]. Although various settings are typically considered, the user is always limited to the arrangements realized by the authors of the given database. For example, the database by [5] considers three geometries of an 8-microphone linear array, with a source 1 and 2 meter from the array center at 13 angles from -90° to 90°, and T_{60} either 160ms, 360ms or 610ms.

Other arrangements are useful for specific case studies, however, they are less useful for detailed analyses of spatial processors, e.g., their robustness to small perturbations in parameters [6], [7]. To analyze such details, a database containing dense measurements of RIRs must be realized, which is a cumbersome task as the measurements must be precisely repeated for different settings. Smaller databases of this kind were realized, e.g., in [8], [9] to support specific analysis. Alternatively, dense sets of RIRs for various microphone positions can be obtained using large arrays; see, e.g., [10].

This paper aims at filling this gap. A new database comprising dense measurements of RIRs is described. The measurements were realized in the acoustic lab at the Bar Ilan University using a device for precise positioning of a loudspeaker, allowing for positioning the source in a dense grid of points within a relatively small volume. Recordings were acquired using six linear arrays of microphones; the measurements were repeated for three different levels of reverberation time.

This work was partly supported by The Czech Science Foundation through Project No. 20-17720S and by the Erasmus+ KA 107 project No. 2017-1-CZ01-KA107-034883; and the Israeli Innovation Authority through KAMIN Project No. 61916, "Environment-Aware Data-Driven Acoustic Signal Processing". Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

[1] https://signalprocessingsociety.org/get-involved/audio-and-acoustic-signal-processing/online-resources

The database provides a new tool for detailed analyses of spatial processing methods, e.g., for source enhancement, localization, and separation. Also, it can be used for learning and testing of new mathematical models of acoustics; field or of RIRs; see, e.g., [11], [12], [13], [14]. The database, dubbed MIRaGe: Multichannel room Impulse Response database on Grid, is available to the research community free-of-charge.

The paper is organized as follows. Section 2 describes the recording setup, the resulting database, and the associated software package that will also be provided. Section 3 describes simple experiments, demonstrating the usage of the database. Section 4 concludes the paper.

II. Database description

A. Setup

The recording setup is situated in the acoustic laboratory, which is a $6 \times 6 \times 2.4$ m rectangular room. A loudspeaker emitting the source is located within a cube-shaped volume of dimensions $46 \times 36 \times 32$ cm (from now referred to as *grid*) as shown in Figures 1 and 2. The positions of the loudspeaker form a grid sampled every 2 cm across the x-axis and y-axis and every 4 cm across the z-axis, so there are $24 \times 19 \times 9 = 4104$ possible source positions (grid vertices) in total. In the following, we will use the Matlab-like notation, for example, [:,:,1] means all vertices in the first horizontal slice of the grid.

In addition, there are 25 source positions located outside of the grid (OOG); nine positions are close to the grid and 16 positions are situated along the walls. The center of the grid (the 5th level), as well as the OOG positions, were positioned at the same height of 115 cm. In the grid positions, the loudspeaker is directed in parallel to the y-axis towards the opposite corner of the room (referred to as 'the front of the grid'). In the OOG positions, the loudspeaker is directed towards the center of the room.

The entire setup is recorded by six static linear microphone arrays and one microphone mounted 2 cm in front of the loudspeaker, which is changing its position with the loudspeaker (microphone 31). Three microphone arrays are placed directly in front of the grid at the distance of 1, 2, and 3 m from the center of the grid. The other three arrays are located at the angle of -45° at the same distances; see Fig. 1. All arrays are directed towards the grid centre and placed at the same height of 115 cm. Each array consists of 5 microphones with the inter-microphone spacing of -13, -5, 0, $+5$ and $+13$ cm relative to the central microphone.

B. Realization

The sidewalls and the ceiling of the acoustic lab consist of revolving double-sided panels with a reflecting and an absorbing face. By rotating these panels, the ratio between the reflective and absorbing areas of the room can be modified, by which the reverberation time of the room can be controlled. We have chosen three reverberation time levels: 100, 300, and 600 ms; T_{60} was measured by Brüel & Kjær type 2250 sound level meter employing Brüel & Kjær Omni source loudspeaker type 4295; see examples of RIR Energy Decay

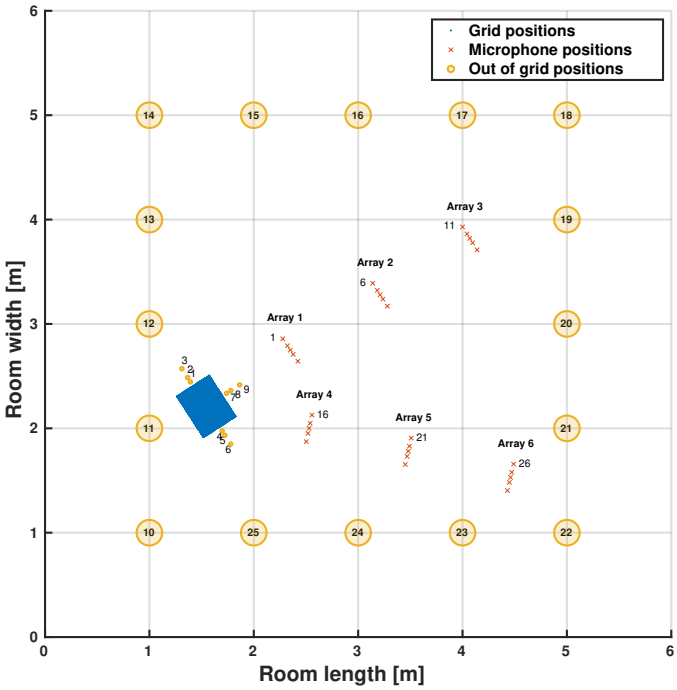

Fig. 1: Illustration of the recording setup - Top View.

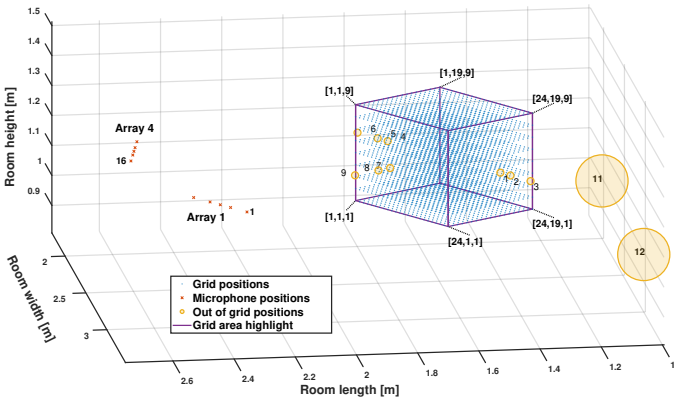

Fig. 2: A detailed view of the near area of the grid.

Curves in Fig. 5. The other details of the recording hardware are summarized in Table I.

The positioning of the loudspeaker within the grid was realized using a precise three-axis positioning system. It consists of a 2D plotter and a lift table. The x and y-axis were controlled automatically by the plotter, while the z-axis was operated manually. The height was measured by a laser distance meter mounted to the desk of the table. To attenuate acoustic reflections from the plotter's rails, a cardboard construction that holds the loudspeaker out of the plotter's perimeter was constructed (see the first photo in Fig. 3). Positions 10 through 25 were measured manually, so the accuracy of their positioning is slightly lower compared to that of the positions in the grid.

For each position, two excitation signals were played and recorded in sequence: The first signal (Chirp) consists of two repetitions of a logarithmic swept-frequency cosine signal with

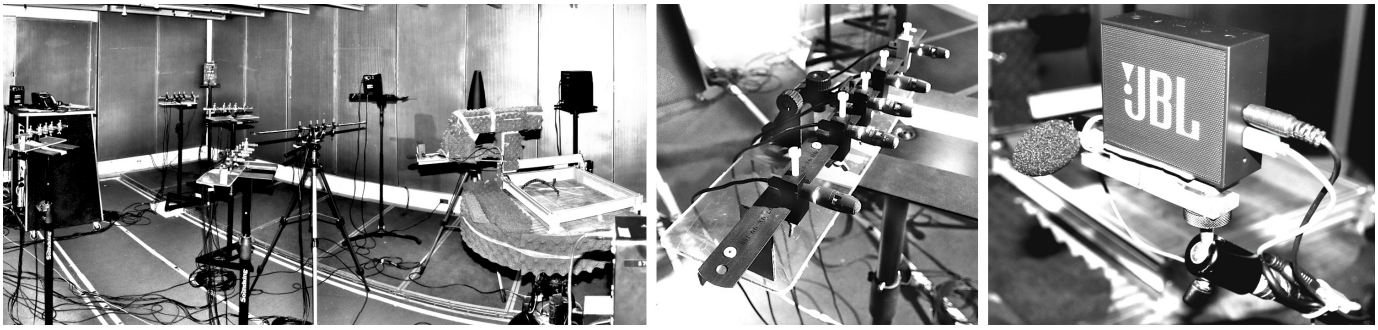

Fig. 3: Photos of the recording setup in the acoustic lab. The first photo from left captures the whole room setup with the 3D positioning system and microphones arrays. The second photo shows the microphones array geometry. The third photo shows the loudspeaker located in an OOG position with the mounted microphone.

TABLE I: Recording equipment

Microphones	AKG CK32
Mic preamp. + AD/DA	ANDIAMO.MC
Loudspeaker (grid, OOG)	JBL Go
Loudspeaker (babble noise)	6301bx Fostex

a total length of 20 seconds (0.5 s silence, 8 s chirp, 2 s silence, 8 s chirp, 1.5 s silence). White Gaussian noise (WN) was used as the second excitation signal with a total length of 10 seconds (0.5s silence, 8s WN, 1.5s silence). Due to the limited frequency range of the loudspeaker (180 Hz - 20 kHz), the Chirp signals starts from 200 Hz through 16 kHz. To prevent rapid phase changes and subsequent "popping", all excitation signals were linearly faded in and out over 0.2 seconds.

We have also recorded one hour of *room tone* (silence) and one hour of diffuse babble noise for each T_{60} setting. The babble noise was simulated by using eight loudspeakers each playing a different multi-speech sequence and each placed approximately 1 m from the walls: one in each corner and one in the middle of each wall. The loudspeakers were directed towards the nearest corner or the nearest wall. Four of these loudspeakers can be seen in the first photo in Fig. 3.

The recording/playback was done with the sampling frequency of 48 kHz and the bit rate of 32 bits per sample. In order to reduce the size of the raw database (1.5 TB), all recordings were re-encoded into the FLAC format (Free Lossless Audio Codec) with 48 kHz and 24 bits per sample (50% reduction).

C. Database package

The database can be downloaded[2] part-by-part according to the directory structure shown in Fig. 4. Each directory at the bottom level in Fig. 4 comprises seven folders: one per each microphone array (01, ..., 06) and one for the microphone mounted to the loudspeaker (on_SPK_mic). A software package for Matlab is available at the same webpage[2]; the software documentation is included. It enables users to compute RIRs and relative RIRs or, equivalently, ATFs or Relative Transfer Functions (RTF) [15] directly from the raw recordings for selected positions, microphones and the other parameters. The entire database has 0.7 TB (FLAC) in size, however, it allows partial download of data to be used with the software package.

[2]https://asap.ite.tul.cz/downloads/MIRaGe

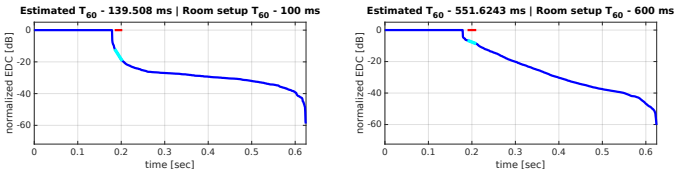

Fig. 5: Samples of RIR Energy Decay Curves (EDCs) created from the MIRaGe database (mic 1, position [12,10,5]); T_{60} was estimated from manually selected (marked - red, cyan) linear parts of EDCs

D. Known bug: Inconsistent System Delay

MIRaGe provides two ways of obtaining a parameterized description of the path between the loudspeaker and microphones. The first is to compute the RIR between on_SPK_mic and any desired array microphone. The second is to compute the RIR between the original excitation signal and the signal recorded by the desired microphone. The estimated impulse response obtained by the second approach includes the response of the audio system (delays of the system and loudspeaker characteristics).

After analyzing the set of RIRs extracted from the grid, we found that the delays between the original excitation signals and the signals recorded by microphones are constant within each XY-plane (within a recording session), however, vary between different XY planes (different recording sessions). This was caused by restarts of the recording hardware and initializations of the ASIO interface through MATLAB between the sessions.

This bug can potentially devalue the extracted RIRs. It can, however, be solved by adding appropriate delay corrections to the excitation signals during the RIR computation. These corrections can be calculated using sounds recorded by the microphone mounted to the loudspeaker and allow us to unify the system delays over all recordings. We have included the delay correction in the software, where it can be turned on or off.

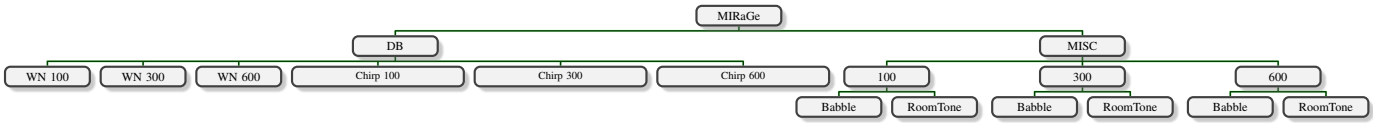

Fig. 4: Database directory structure: The acronyms WN and Chirp correspond to the excitation signals as defined in Section II-B. The numbers 100, 300, and 600 correspond to the reverberation time level. MISC contains the recordings of the room tone and of the babble noise.

III. Practical Use

A. Spatial Map of Signal-to-Interference Ratio

In this example, MIRaGe is used for simulating an acoustic scenario and a subsequent detailed analysis of a beamformer. A target source at position in [12,10,5] is emitting a chirp source. Two inteference sources (WN) are placed in [3, 10, 5] and [22, 10, 5]. Signal images as received by microphones of Array 1 are first generated. Then, an MVDR beamformer focused on the target source is computed using the known source images with microphone 1 as a reference.

Now, the goal is to visualize the spatial map of Signal-to-Interference Ratio (SIR) improvement measured at the output of the MVDR. To this end, the beamformer's weights are fixed while the location of the chirp source is considered in all positions within the grid. The SIR improvement, as a function of the chirp source position, corresponds to the desired map. Fig. 6 shows such maps abtained for the three T_{60} settings available in MIRaGe.

Obviously, the maps show the highest SIR improvement in the positions focused by the beamformer, and the lowest values are in the positions of the interferers. While the beam directed towards the focused position is rather wide in terms of angle and distance from the microphones, the null beams directed towards the interferers are narrow and concentrated on the particular positions. This suggests that MVDR is spatially robust against the true target's position, but more sensitive to the positions of interferers.

More details of this experiment can be seen through a MATLAB application that is available on the database webpage. The application enables the user to browse more settings: changing the focused position of the beamformer, comparing maps obtained through MIRaGe and the RIR generator [16], switching between MVDR/MPDR beamformers, showing maps of more metrics, and doing octave-band analyses.

B. Manifold Learning

In [11], a manifold projection (MP) method for supervised RTF identification was introduced for the purpose of increasing the robustness of optimal beamformers. In this method, a manifold of typical RTFs in a particular room is learned in advance and then exploited to improve the RTF estimation from noisy measurements. So far, this method was tested only using simulated experiments [11], where it was shown to provide superior results, especially, in low SNR conditions. Here, we provide an evaluation using our database.

To repeat the experimental setup of [11] as closely as possible, we use every second position in the grid's x-axis,

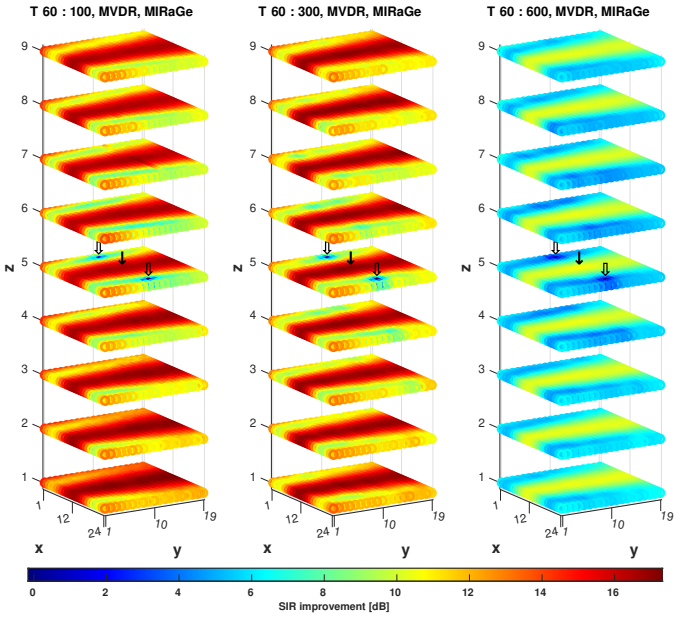

Fig. 6: Maps of the SIR improvement of the MVDR beamformers in three T_{60} settings; ↓ points to the focused (target) position of the beamformers; ⇓ point to the positions of interferers.

every position of the z-axis and the first position of the y-axis (that is, [1:2:24,1,:]). For these positions, the RTFs are calculated between microphones 14 and 15 (Array 3, 8 cm inter-mic distance, 3 m in front of the grid) using clean excitation white Gaussian noise signals. These RTFs serve as the training set.

In each trial, a random position within the grid except any positions in [:,1,:] is selected. A 3-second long speech signal is simulated from that position. Then, an uncorrelated spatial white Gaussian noise is added at different signal-to-noise ratio (SNR) levels. From the noisy mixture, an RTF is estimated using the frequency-domain estimator from [15]. This estimate is compared to the Nearest Neighbor (NN) RTF and to the enhanced estimate by the MP. The NN approach is taking the RTF from the training set that is closest to the estimate in terms of the Euclidean distance. As a criterion, the blocking ability with respect to the simulated source is computed as follows: Given an RTF estimate, the blocking ability for microphones i and j is

$$\mathrm{BA}_{i,j}(g) = -10\log_{10}\left(\frac{\mathrm{var}[s_r]/\mathrm{var}[v_r]}{\mathrm{var}[s]/\mathrm{var}[v]}\right) \qquad (1)$$

where s_r is the residual of the source signal s at the output

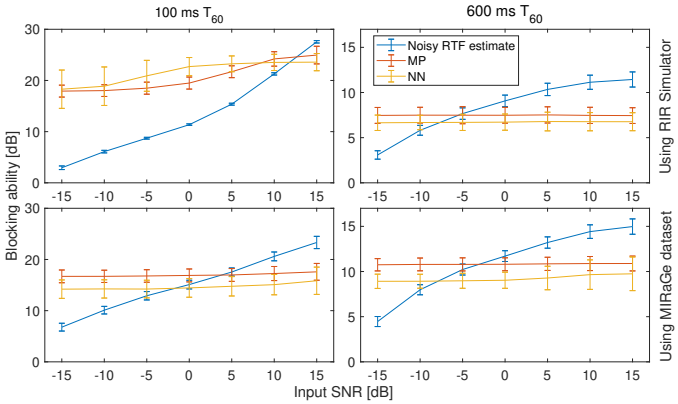

Fig. 7: Comparison of RTF estimation methods in terms of blocking ability. Reverberations of 100 ms T_{60} (left), 600 ms T_{60} (right), using RIR simulator (top) and MIRaGe dataset (bottom).

of the blocking operation $x_i * g - x_j$. Here, x_i and x_j are the signals on the microphones, and g is the estimated ReIR (the time-domain counterpart of the RTF estimate). Similarly, v_r is the residual of the noise. Once g is the true ReIR, then $s_r = 0$ and $\mathrm{BA}_{i,j}(g) = +\infty$.

Averaged results of 1000 independent trials are shown in Fig. 7. The results show that both supervised methods provide superior blocking ability compared to the unsupervised method (the noisy RTF estimate) when the input SNR is lower than 0 dB or -10 dB, depending on the reverberation time. While both NN and MP methods perform similarly in the simulated conditions and low reverberation time, MP provides consistently better results than NN in terms of the blocking ability when using real data.

IV. Conclusions

We have introduced a new database of dense measurements within a 3D area of an acoustic lab. We have shown that the database can be used for detailed analyses of spatial processing algorithms subject to source location within the measured area. Small variations of RIRs and ReIRs (resp. ATFs and RTFs) due to small changes of the source location can be observed. The database provides an alternative to the popular room impulse response simulator based on the image method [1], [16].

References

[1] Jont B. Allen and David A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[2] M. Jeub, M. Schafer, and P. Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *2009 16th International Conference on Digital Signal Processing*, July 2009, pp. 1–5.

[3] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. 1, pp. 298605, Jul 2009.

[4] R. Stewart and M. Sandler, "Database of omnidirectional and b-format room impulse responses," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2010, pp. 165–168.

[5] E. Hadad, F. Heese, P. Vary, and S. Gannot, "Multichannel audio database in various acoustic environments," in *2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sep. 2014, pp. 313–317.

[6] Y. R. Zheng, R. A. Goubran, and M. El-Tanany, "Robust near-field adaptive beamforming with distance discrimination," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 478–488, Sep. 2004.

[7] A. Barnov, V. B. Bracha, S. Markovich-Golan, and S. Gannot, "Spatially robust gsc beamforming with controlled white noise gain," in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, Sep. 2018, pp. 231–235.

[8] M. Fakhry and F. Nesta, "Underdetermined source detection and separation using a normalized multichannel spatial dictionary," in *IWAENC 2012; International Workshop on Acoustic Signal Enhancement*, Sep. 2012, pp. 1–4.

[9] Z. Koldovský, J. Málek, P. Tichavský, and F. Nesta, "Semi-blind noise extraction using partially known position of the target source," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 10, pp. 2029–2041, Oct 2013.

[10] Eugene Weinstein, Kenneth Steele, Anant Agarwal, and James Glass, "LOUD: A 1020-node microphone array and acoustic beamformer," in *14th International Congress on Sound and Vibration (ICSV)*, July 2007.

[11] Ronen Talmon and Sharon Gannot, "Relative transfer function identification on manifolds for supervised GSC beamformers," in *21st European Signal Processing Conference (EUSIPCO 2013)*. IEEE, 2013, pp. 1–5.

[12] Bracha Laufer-Goldshtein, Ronen Talmon, and Sharon Gannot, "A study on manifolds of acoustic responses," in *Latent Variable Analysis and Signal Separation*, Emmanuel Vincent, Arie Yeredor, Zbyněk Koldovský, and Petr Tichavský, Eds., Cham, 2015, pp. 203–210, Springer International Publishing.

[13] B. Laufer-Goldshtein, R. Talmon, and S. Gannot, "A real-life experimental study on semi-supervised source localization based on manifold regularization," in *2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE)*, Nov 2016, pp. 1–5.

[14] F. Katzberg, R. Mazur, M. Maass, P. Koch, and A. Mertins, "A compressed sensing framework for dynamic sound-field measurements," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 11, pp. 1962–1975, Nov 2018.

[15] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1614–1626, Aug 2001.

[16] Emanuël A.P. Habets, "Room impulse response generator," *Technische Universiteit Eindhoven, Tech. Rep*, vol. 2, no. 2.4, 2006.