BRUDEX Database: *B*inaural *R*oom Impulse Responses with *U*niformly *D*istributed *Ex*ternal Microphones

Daniel Fejgin*, Wiebke Middelberg*, Simon Doclo

University of Oldenburg, Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4all, Oldenburg, Germany
Email: {daniel.fejgin,wiebke.middelberg,simon.doclo}@uni-oldenburg.de
Web: www.sigproc.uni-oldenburg.de

Abstract

There is an emerging need for comparable data for multi-microphone processing, particularly in acoustic sensor networks. However, commonly available databases are often limited in the spatial diversity of the microphones or only allow for particular signal processing tasks. In this paper, we present a database of acoustic impulse responses and recordings for a binaural hearing aid setup, 36 spatially distributed microphones spanning a uniform grid of (5×5) m^2 and 12 source positions. This database can be used for a variety of signal processing tasks, such as (multi-microphone) noise reduction, source localization, and dereverberation, as the measurements were performed using the same setup for three different reverberation conditions ($T_{60} \approx \{310, 510, 1300\}$ ms). The usability of the database is demonstrated for a noise reduction task using a minimum variance distortionless response beamformer based on relative transfer functions, exploiting the availability of spatially distributed microphones.

1 Introduction

Signal processing for acoustic sensor networks is a field of increasing interest [1–6] as spatially distributed microphones allow for a more diverse spatial sampling of the sound field than compact microphone arrays. To enable reproducible and comparable research, publicly available databases that allow for different signal processing tasks, such as speech enhancement or sound source localization, are of utmost importance. Despite the availability of simulation methods and tools for the generation of room impulse responses (RIRs), e.g., [7–9], measured RIRs and real-world recordings are indispensable to evaluate the performance of algorithms. Nowadays, there is a variety of multi-microphone RIR databases available, e.g., considering linear arrays [10], widely distributed microphones [11, 12], head-mounted devices like hearing aids [13, 14], and head-mounted or body-worn microphones together with distributed microphones [15–17].

In this paper, we present a new, complementary database, referred to as **B**inaural **R**oom Impulse Responses with **U**niformly **D**istributed **Ex**ternal microphones (BRUDEX). It consists of a total of about 1500 measured RIRs for three different reverberation conditions, 12 source positions, binaural hearing aids on a dummy head (four hearing aid microphones and two in-ear microphones), and spatially distributed microphones at 36 uniformly distributed positions (see Figure 1). To distinguish between the head-mounted microphones (i.e., hearing aid and in-ear microphones) and the spatially distributed microphones, we will refer to the latter as external microphones (eMics). The database was recorded in an acoustic laboratory at the University of Oldenburg (see Figure 2). The reverberation condition in the laboratory can be set by means of curtains and absorber panels, which are mounted to the walls and the ceiling. The RIRs were measured for three reverberation conditions, keeping the microphone and loudspeaker configuration unchanged while varying the reverberation condition. Besides measured RIRs, the database contains recordings of speech and noise signals for all mentioned conditions. A summary of the content of the database is presented in Table 1. The BRUDEX database allows for a variety
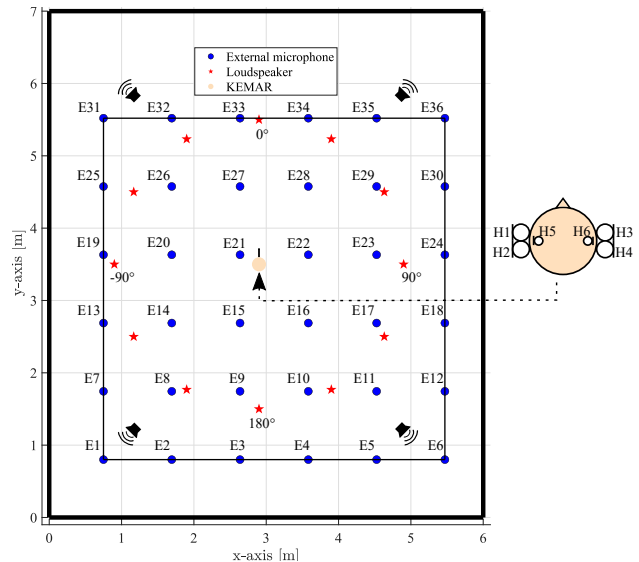
Figure 1: Placement of external microphones (E1-E36, ●), loudspeakers (★), KEMAR dummy head with four hearing aid microphones (H1-H4) and two in-ear microphones (H5-H6, see close-up), and four noise loudspeakers (◀») to generate a quasi-diffuse noise field in a laboratory with variable reverberation.

of signal processing tasks using spatially distributed microphones: First, the database obviously allows for multi-microphone noise reduction and speech enhancement, as both measured RIRs as well as separate speech and noise recordings are available. Second, the database allows for dereverberation in different acoustic conditions, as measurements for different reverberation conditions are available. Third, the database allows for source localization, as all microphone and loudspeaker positions are calibrated.

In Sections 2 and 3, we provide details on the setup, the measurement conditions and the calibration and measurement procedures. In Section 4, we provide practical information about the accessibility. To demonstrate the usability of the BRUDEX database and the compatibility with the database in [13], in Section 5 we construct an acoustic scenario consisting of a target speech source and diffuse-like background noise using recordings from the four hearing aid microphones and three external microphones. We consider a binaural minimum variance distortionless response beamformer steered by the relative transfer function vector [18], which is either estimated blindly from the microphone signals or obtained from database RIRs. The simulation results show a good accordance with the results from recent literature.

2 Setup and Spatial Configurations

In this section, we provide an overview of the acoustic setup and the calibration procedure for the BRUDEX database. In Section 2.1, we describe the laboratory and its acoustic properties and in Section 2.2, we describe the configuration of the loudspeakers and microphones.

Table 1: Content of the BRUDEX database.

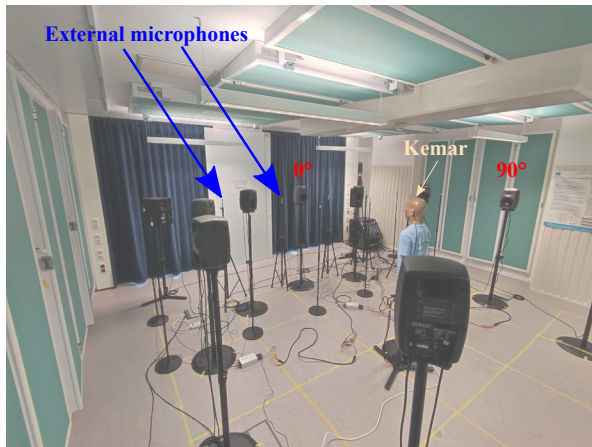|	RIR	speech signal				noise signal		
Reverb condition	low, medium, high	low, medium, high				low, medium, high		
Type	–	female	female	male	male	babble	cafeteria	white
Duration [s]	2	20	30	20	30	20		
DOA [°]	-150:30:180	-150:30:180				quasi-diffuse		

Figure 2: *Variable Acoustics Laboratory in the low reverberation condition, with all absorber panels open and curtains closed. The picture shows external microphones located on the positions E25-E36 (see Figure 1).*

Figure 3: Violin plots of the distribution of the reverberation time T_{60} (left) and the direct-to-reverberation ratio (right) for the three considered reverberation conditions. Median values are shown on top of the figure.

2.1 Acoustic Properties of the Laboratory

All measurements were performed in the *Variable Acoustics Laboratory* at the University of Oldenburg that has dimensions of about $(7 \times 6 \times 2.7)\,\text{m}^3$ (see Figure 2). The measurements were performed for three reverberation conditions referred to as low, medium and high. To quantify the amount of reverberation, we consider the reverberation time T_{60} as well as the direct-to-reverberation ratio (DRR). The reverberation time of each RIR was determined as the extrapolated decay time from -5dB to -35dB on the logarithmic energy decay curve, obtained using Schroeder's backward integration method [19]. The DRR of each RIR was obtained as the ratio of the power of the direct-path component and the power of the early and late reflections [20]. We define the duration of the direct-path component to be 8 ms larger than the approximate arrival time of the direct sound. For each reverberation condition, Figure 3 depicts violin plots of the distribution of T_{60} and DRR, obtained from the measured RIRs of all loudspeaker-microphone pairs (see Sections 2.2 and 3). For each reverberation condition, the distributions of T_{60} and DRR clearly show strong variations for different loudspeaker-microphone pairs, mainly depending on their distance and their positions relative to the walls. The median values for these reverberation parameters are $T_{60} \approx \{310, 510, 1300\}$ ms and DRR $\approx \{3.5, -0.5, -4.0\}$ dB.

2.2 Microphone and Loudspeaker Configuration

As mentioned above, Figure 1 depicts the positions of the microphones, the loudspeakers, and the dummy head. We considered head-mounted microphones as well as spatially distributed microphones, which we refer to as eMics. The six head-mounted microphones consist of four microphones of a binaural hearing aid (two microphones on each hearing aid with an inter-microphone distance of about 15 mm) mounted on a G.R.A.S. KEMAR 45 BM dummy head with anthropometric pinnae and two low noise in-ear
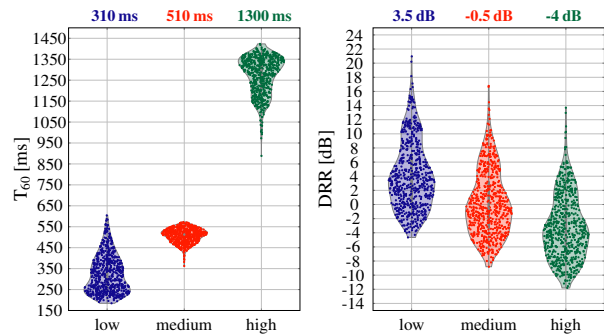
microphones. It should be noted that the used hearing aids are the same as the hearing aids in the database [13], where we considered the front and rear microphones. The dummy head was located approximately in the center of the laboratory, with its ears at an approximate height of 1.5 m.

Besides the head-mounted microphones, a total of 36 eMic positions were considered in this database. Omnidirectional microphones (type: Sennheiser KE 4-211-2) were placed on a uniform grid of about $(5 \times 5)\,\text{m}^2$ around the dummy head with an inter-microphone spacing of (95 ± 2)cm in x- and y-direction at a height of about 1.5 m. The positions of the eMics were manually calibrated by means of a cross-line laser and a laser distance meter, resulting in a positioning error of about 2 cm.

For the measurement of the RIRs and the recording of the speech signals, 12 loudspeakers (types GENELEC 8030A, 8030B and 8330 APM) were placed at a distance of about (200 ± 2)cm at angles $(-150.0\pm2.5)°$, $(-120.0\pm2.5)°$,...,$(180.0\pm2.5)°$, relative to the look-direction of the dummy head. Each loudspeaker was set to a height of approximately 1.5m. The distance between the loudspeakers and the dummy head was verified using a laser distance meter. Based on three procedures with a resolution of $5°$, the angles between the loudspeakers relative to the dummy head were verified [21–23].

For the recording of the noise signals, four loudspeakers (M-Audio BX8 D2) were placed at about 170 cm from (and facing) the corners of the room.

3 Measurement Procedure

All measurements were performed at a sampling rate of 48 kHz using an RME MADIface XT audio interface and RME ADI-8 QS and RME OctaMic XTC analog-digital-converters. For all recordings, the input-output delay due to the buffering of the audio interface was compensated. The transfer functions of the loudspeakers and of the microphones were not compensated. For the RIR measurements, we used the exponential sine sweep technique [24] using synchronized swept sine signals [25] as excitation signals. The excitation signal had a duration of about 5 s and the frequency increased exponentially in the range $20\,\text{Hz} - 24\,\text{kHz}$. Playing back the excitation signal with each of the loudspeakers separately, the excitation signal was recorded 10 times with all

available microphones simultaneously and averaged. By convolving this average recording with the inverse sweep and considering only the linear part of the resulting set of higher-order impulse responses, the RIRs between the loudspeaker and the microphones were obtained. To smoothen and truncate the RIRs to a duration of 2 s, Hanning windows with a duration of 0.38 ms for the fade-in and 50 ms for the fade-out were applied.

As only 12 identical microphones were available for the spatially distributed microphones, for each reverberation condition we performed three separate recordings to cover all eMic positions E1 - E36. We refer to each separate recording as a run. We only changed the position of these 12 eMics after performing all measurements for all reverberation conditions (since changing the reverberation condition is more reproducible than changing the microphone position). In each run we recorded 18 channels simultaneously, i.e., the six head-mounted microphones (H1-H6, at the same position for all runs) and 12 eMics (run 1: positions E1 - E12, run 2: positions E13 - E24, and run 3: positions E25 - E36). Since due to scattering effects the positioning of the eMics may affect the signals recorded at the head-mounted microphones, the database provides two types of RIRs for the head-mounted microphones, which we refer to as individual and average. For the individual head-mounted RIRs the measured responses were considered for each run individually. For the average head-mounted RIRs the measured recordings were averaged over the three runs, (i.e. run-averaged) before performing the convolution with the inverse sweep.

The multi-channel speech recordings correspond to two female and two male English speech signals (duration: 20 or 30 s) from the SQAM and VCTK corpus [26, 27]. Similarly as for the RIRs, each speech signal was played back with each of the loudspeakers separately, and recorded with all available microphones simultaneously. The multi-channel noise recordings correspond to babble noise, cafeteria noise, and white noise (duration: 20 s). The noise was generated with the four loudspeakers facing the corners of the laboratory, playing back different (uncorrelated) versions of the respective noise. Similarly as for the RIRs, for the speech and noise recordings we also distinguish the individual and run-averaged recordings for the head-mounted microphones.

4 Database

The BRUDEX database is available under https://doi.org/10.5281/zenodo.7986446 under an MIT License. The measured RIRs and the speech and noise recordings are located in sub-directories organized by the reverberation condition. The file names encode the DOA of the used loudspeaker (not applicable for noise recordings) and indicate whether the head-mounted microphone signals are obtained as individual recordings (18 channel signals) or as run-averaged recordings (42 channel signals). For the former case, the files contain 18 channels corresponding to the six head-mounted microphones and the 12 eMics. For the latter case, the files contain 42 channels corresponding to the six head-mounted microphones and the eMics at the 36 positions. The content of the database is organized in separate uncompressed binary MAT-files. To facilitate usage of the database, we provide MATLAB and Python scripts for accessing the multi-channel data.

5 Application to MVDR Beamforming

To demonstrate the usability of the BRUDEX database, in this section we consider an exemplary multi-microphone noise reduction application. More particularly, we consider a binaural minimum variance distortionless response (MVDR) beamformer using the hearing aid microphones and three eMics. Besides using estimated relative transfer function vectors to steer the MVDR beamformer, we furthermore compare the performance of different database RIRs to investigate the robustness against non-matching RIRs. We briefly review the used algorithms in Section 5.1, the considered acoustic scenario and the implementation in Section 5.2, and the simulation results in Section 5.3.

5.1 Algorithms

We consider an acoustic scenario with a single target speaker and quasi-diffuse background noise and M microphones composed of $M_{\mathrm{H}} = 4$ hearing aid microphones and $M_{\mathrm{E}} = 3$ eMics. In the short-time Fourier transform (STFT) domain, with k denoting the frequency bin index and l denoting the frame index, the M-dimensional noisy signal vector $\mathbf{y}(k,l)$ can be written as

$$\mathbf{y}(k,l) = \mathbf{x}(k,l) + \mathbf{n}(k,l) = \mathbf{h}(k,l)X_{\mathrm{ref}}(k,l) + \mathbf{n}(k,l), \quad (1)$$

where $\mathbf{x}(k,l)$ and $\mathbf{n}(k,l)$ denote the speech and noise component, respectively. We assume a multiplicative transfer function model [28] such that the speech component can be written in terms of its relative transfer function (RTF) vector $\mathbf{h}(k,l)$, which relates the speech component in a reference microphone $X_{\mathrm{ref}}(k,l)$ to the speech components in all other microphones. In the following, the indices k and l are omitted wherever possible. By assuming uncorrelated speech and noise components, the noisy covariance matrix $\mathbf{R}_y = \mathcal{E}\{\mathbf{y}\mathbf{y}^H\}$ can be decomposed into the speech covariance matrix $\mathbf{R}_x$ and the noise covariance matrix $\mathbf{R}_n$, i.e.,

$$\mathbf{R}_y = \mathbf{R}_x + \mathbf{R}_n, \quad (2)$$

where $\mathcal{E}\{\cdot\}$ denotes the expectation operator and $\{\cdot\}^H$ denotes the Hermitian transpose operator. The MVDR beamformer [29–31] aims at minimizing the noise power spectral density while preserving the speech component in the reference microphone. It was shown in [29, 30] that the MVDR beamformer is given by

$$\mathbf{w} = \frac{\hat{\mathbf{R}}_n^{-1}\hat{\mathbf{h}}}{\hat{\mathbf{h}}^H \hat{\mathbf{R}}_n^{-1}\hat{\mathbf{h}}}, \quad (3)$$

which requires estimates of the RTF vector $\hat{\mathbf{h}}$ and the noise covariance matrix $\hat{\mathbf{R}}_n$.

To estimate the RTF vector for the MVDR beamformer in (3), in [32] and [6] methods were proposed that exploit one or multiple eMics, assuming that the noise component in each eMic signal is uncorrelated with the noise components in all other microphone signals. This assumption holds quite well, e.g., for a diffuse noise field where the eMics are spatially separated from each other and from the hearing aid microphones [32]. The so-called spatial coherence (SC) method proposed in [32] estimates the RTF vector using the m_{E}-th eMic by simply selecting the column of the noisy covariance matrix corresponding to the m_{E}-th eMic and dividing it by the reference entry, i.e.,

$$\hat{\mathbf{h}}_{m_{\mathrm{E}}}^{\mathrm{SC}} = \frac{\hat{\mathbf{R}}_y \mathbf{e}_{m_{\mathrm{E}}}}{\mathbf{e}_{\mathrm{ref}}^T \hat{\mathbf{R}}_y \mathbf{e}_{m_{\mathrm{E}}}} \quad (4)$$

where $\mathbf{e}_{m_{\mathrm{E}}}$ denotes a selection vector for the m_{E}-th eMic and $\mathbf{e}_{\mathrm{ref}}$ denotes a selection vector for the reference microphone(s), i.e., one for the left and right hearing aid each to allow for binaural processing. Since for each of the M_{E} available eMics a different RTF vector estimate $\hat{\mathbf{h}}_{m_{\mathrm{E}}}^{\mathrm{SC}}$ is obtained, it was proposed in [6] to linearly combine them as $\hat{\mathbf{h}} = \hat{\mathbf{H}}\boldsymbol{\alpha}$, where $\hat{\mathbf{H}}$ denotes a matrix containing all M_{E} estimates, i.e.,

$$\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1^{\mathrm{SC}},...,\hat{\mathbf{h}}_{M_{\mathrm{E}}}^{\mathrm{SC}}], \quad (5)$$

and $\boldsymbol{\alpha}$ denotes a weight vector. A particular linear combination proposed in [6] aims at maximizing the output signal-to-noise ratio (SNR) of the MVDR beamformer by means of the weight vector $\boldsymbol{\alpha}^{\mathrm{mSNR}}$. The weight vector of the so-called mSNR method is given by [6]

$$\boldsymbol{\alpha}^{\mathrm{mSNR}} = \frac{\mathcal{P}\{\mathbf{B}^{-1}\mathbf{A}\}}{\mathbf{1}_{M_{\mathrm{E}}\times 1}^T \mathcal{P}\{\mathbf{B}^{-1}\mathbf{A}\}}, \quad (6)$$

where $\mathcal{P}\{\cdot\}$ denotes the principal eigenvector operator, $\mathbf{1}_{M_{\mathrm{E}}\times 1}$ denotes an M_{E}-dimensional vector of ones and $\mathbf{B} = \hat{\mathbf{H}}^H \hat{\mathbf{R}}_n^{-1}\hat{\mathbf{H}}$ and $\mathbf{A} = \hat{\mathbf{H}}^H \hat{\mathbf{R}}_n^{-1}\hat{\mathbf{R}}_y \hat{\mathbf{R}}_n^{-1}\hat{\mathbf{H}}$. The RTF vector estimate using the mSNR method is obtained as $\hat{\mathbf{h}}^{\mathrm{mSNR}} = \hat{\mathbf{H}}\boldsymbol{\alpha}^{\mathrm{mSNR}}$. For further details, we refer the reader to the literature mentioned above.

5.2 Acoustic Scenario and Implementation

To evaluate the performance of different MVDR beamformers, we consider an acoustic scenario with a female target speaker (SQAM) at $-60°$ relative to the dummy head and quasi-diffuse babble noise in the medium reverberation condition. In addition to the four hearing aid microphones (H1-H4), we used three eMics at the positions E14, E27 and E28. For the hearing aid microphones, we used the run-averaged RIRs and recordings.

All microphone signals were generated by using the recorded speech and noise components from the database at a broadband input SNR of 0 dB in the front microphone on the left hearing aid by scaling the noise components accordingly. The resulting input SNR in the front microphone on the right hearing aid was -1 dB and in the eMics about 0 dB to 1 dB.

The processing was implemented in the STFT domain using a square-root-Hann window for analysis and synthesis with a frame length of 2048 samples (corresponding to about 42 ms at a sampling rate of 48 kHz) and 50 % overlap. The covariance matrices were estimated using recursive smoothing, i.e.,

$$\hat{\mathbf{R}}_y(k,l) = \alpha_y \hat{\mathbf{R}}_y(k,l-1) + (1-\alpha_y)\mathbf{y}(k,l)\mathbf{y}^H(k,l), \quad (7)$$

$$\hat{\mathbf{R}}_n(k,l) = \alpha_n \hat{\mathbf{R}}_n(k,l-1) + (1-\alpha_n)\mathbf{y}(k,l)\mathbf{y}^H(k,l), \quad (8)$$

where the parameters α_y and α_n correspond to smoothing times of 500 ms and 1000 ms, respectively. The noisy covariance matrix $\hat{\mathbf{R}}_y$ was estimated during speech activity, i.e., if $\overline{\text{SPP}}(k,l) > 0.5$, whereas the noise covariance matrix $\hat{\mathbf{R}}_n$ was estimated during speech pauses, i.e., if $\overline{\text{SPP}}(k,l) < 0.5$. $\overline{\text{SPP}}$ denotes the estimated speech presence probability [33] averaged over all eMics.

In the evaluation, we compare the individual RTF vector estimates obtained from the SC method in (4) (denoted as SC-E14, SC-E27 and SC-E28, respectively) and the mSNR estimate obtained using the weight vector in (6). To investigate the robustness against non-matching RTFs, we additionally consider the RTF vector obtained from the following database RIRs:

- RIR (medium): RIRs for all used microphones for DOA = $-60°$ in the medium reverberation condition, i.e., matching the generated microphone signals.
- RIR (low): RIRs for all used microphones for DOA = $-60°$ in the low reverberation condition, i.e., not matching the generated microphone signals.
- RIR (anechoic): anechoic RIRs from [13] for the hearing aid microphones for DOA = $-60°$, i.e., an approximation with no reverberation. It should be noted that in this case only the RIRs of the hearing aid microphones and not for the eMics are available, such that the MVDR beamforming is only performed using the hearing aid microphones.

The RTFs were computed from the RIRs by convolving the RIRs with white Gaussian noise and computing the principal eigenvector of the resulting covariance matrix in the STFT domain.

To assess the performance of the different beamformers, we consider the broadband binaural SNR improvement (ΔBSNR) of the binaural MVDR beamformer compared to the unprocessed noisy reference microphone signals. The SNRs are computed in overlapping segments of 1 s length (overlapping by 85%) during speech activity and averaged over time. To allow for a suitable initialization of the covariance matrices, the first 1.5 s of the signal are not taken into account in the computation of the SNR.

5.3 Simulation Results

Figure 4 depicts the BSNR improvement (along with the variance as error bars) for the different considered RTF vectors.

For the RTF vectors computed from database RIRs, i.e., RIR (anechoic), RIR (low) and RIR (medium), the following observations can be made: Using the RIRs from the matching reverberation condition yields the best performance (about 7 dB improvement in BSNR), while using the RIRs from the low reverberation condition still yields a rather good performance (about 4 dB improvement in BSNR). Yet, the performance decreases compared to the true RIRs (RIR (medium)), as RIR (low) does not account for
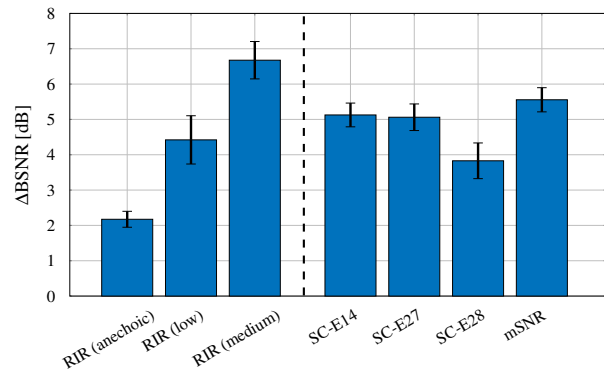

Figure 4: Improvement of the BSNR for the RTF-vector-steered binaural MVDR beamformer using different RTF vector estimation methods. Left: computed from database RIRs, right: estimated from microphone signals.

all reflections. Using the anechoic RIRs yields a rather low BSNR improvement of only 2 dB. This can be explained by the fact that on the one hand only the head-mounted microphones are used in the beamformer and on the other hand by the limitation of the anechoic RIRs to only approximate the direct path but no reflections.

Using the RTF vectors estimated with the SC method (SC-E14, SC-E27 and SC-E28), results in a BSNR improvement of about 4–5 dB. The mSNR approach yields a BSNR improvement of about 5.5 dB, slightly but consistently outperforming the single SC-based RTF vector estimates, but yielding a lower BSNR improvement than using the true RIRs (RIR (medium)). These results correspond well with prior results from literature [6, 34], which indicates a good accordance of the BRUDEX database with established simulation paradigms.

6 Summary

In this paper, we presented the BRUDEX database, a novel database of RIRs and speech and noise recordings for binaural hearing aids and uniformly spaced distributed microphones. We presented the used measurement and calibration procedures and characterized the resulting RIRs in terms of T_{60} and DRR. The applicability of the database was demonstrated by means of a binaural MVDR beamformer where the underlying acoustic scenario was generated from the database. The RTF vectors for steering the beamformer were either estimated using two recently proposed methods that exploit external microphones or using RIRs from databases. The BRUDEX database can be used for many other signal processing tasks such as source localization, dereverberation, and noise reduction.

References

[1] S. Markovich-Golan, A. Bertrand, M. Moonen, and S. Gannot, "Optimal distributed minimum-variance beamforming approaches for speech enhancement in wireless acoustic sensor networks," *Signal Processing*, vol. 107, pp. 4–20, Feb. 2015.

[2] V. M. Tavakoli, J. R. Jensen, M. G. Christensen, and J. Benesty, "A framework for speech enhancement with ad hoc microphone arrays," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 24, pp. 1038–1051, Mar. 2016.

[3] M. Cobos, F. Antonacci, A. Alexandridis, A. Mouchtaris, and B. Lee, "A survey of sound source localization methods in wireless acoustic sensor networks," *Wireless Communications and Mobile Computing*, vol. 2017, pp. 1–24, Aug. 2017.

[4] A. I. Koutrouvelis, T. W. Sherson, R. Heusdens, and R. C. Hendriks, "A low-cost robust distributed linearly constrained beamformer for wireless acoustic sensor networks with arbitrary topology," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 26, pp. 1434–1448, Apr. 2018.

[5] J. Zhang, R. Heusdens, and R. C. Hendriks, "Relative acoustic transfer function estimation in wireless acoustic sensor networks," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 27, pp. 1507–1519, Jun. 2019.

[6] N. Gößling, W. Middelberg, and S. Doclo, "RTF-steered binaural MVDR beamforming incorporating multiple external microphones," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, (New Paltz, NY, USA), pp. 368–372, Oct. 2019.

[7] D. P. Jarrett, E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor, "SMIR Generator," 2019. [Online]. Available: https://www.audiolabs-erlangen.de/fau/professor/habets/software/smir-generator.

[8] T. Wendt, S. van de Par, and S. D. Ewert, "A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation," *Journal of the Audio Engineering Society*, vol. 62, no. 11, pp. 748–766, 2014.

[9] E. De Sena, N. Antonello, M. Moonen, and T. van Waterschoot, "On the modeling of rectangular geometries in room acoustic simulations," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 4, pp. 774–786, 2015.

[10] E. Hadad, F. Heese, P. Vary, and S. Gannot, "Multichannel audio database in various acoustic environments," in *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, (Juan-les-Pins, France), pp. 313–317, Sep. 2014.

[11] R. Stewart and M. Sandler, "Database of omnidirectional and B-format room impulse responses," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Dallas, TX, USA), pp. 165–168, Mar. 2010.

[12] S. Koyama, T. Nishida, K. Kimura, T. Abe, N. Ueno, and J. Brunnström, "MeshRIR: A dataset of room impulse responses on meshed grid points for evaluating sound field analysis and synthesis methods," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, (New Paltz, NY, USA), pp. 1–5, Oct. 2021.

[13] 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, pp. 1–10, Jul. 2009.

[14] F. Denk, S. M. Ernst, S. D. Ewert, and B. Kollmeier, "Adapting hearing devices to the individual ear acoustics: Database and target response correction functions for various device styles," *Trends in hearing*, vol. 22, p. 2331216518779313, Jun. 2018.

[15] W. S. Woods, E. Hadad, I. Merks, B. Xu, S. Gannot, and T. Zhang, "A real-world recording database for ad hoc microphone arrays," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, (New Paltz, NY, USA), pp. 1–5, Oct. 2015.

[16] R. M. Corey, M. D. Skarha, and A. C. Singer, "Massive distributed microphone array dataset," 2019. [Online]. Available: https://doi.org/10.13012/B2IDB-6216881_V1.

[17] T. Dietzen, R. Ali, M. Taseska, and T. van Waterschoot, "MYRiAD: a multi-array room acoustic database," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2023, pp. 1–14, Apr. 2023.

[18] S. Doclo, S. Gannot, D. Marquardt, and E. Hadad, "Binaural speech processing with application to hearing devices," in *Audio Source Separation and Speech Enhancement* (E. Vincent, T. Virtanen, and S. Gannot, eds.), pp. 413–442, Hoboken, NJ, USA: John Wiley & Sons, Ltd., 2018.

[19] M. R. Schroeder, "New method of measuring reverberation time," *The Journal of the Acoustical Society of America*, vol. 37, pp. 419–412, Jun. 1965.

[20] P. A. Naylor, E. A. P. Habets, J. Y.-C. Wen, and N. D. Gaubitch, "Models, measurement and evaluation," in *Speech Dereverberation* (P. A. Naylor and N. D. Gaubitch, eds.), pp. 21–56, London, Great Britain: Springer, 2010.

[21] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. on Antennas and Propagation*, vol. 34, pp. 276–280, Mar. 1986.

[22] H. Kayser and J. Anemüller, "A discriminative learning approach to probabilistic acoustic source localization," in *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, (Juan-les-Pins, France), pp. 99–103, Sep. 2014.

[23] D. Fejgin and S. Doclo, "Comparison of binaural RTF-vector-based direction of arrival estimation methods exploiting an external microphone," in *Proc. European Signal Processing Conference (EUSIPCO)*, (Dublin, Ireland), pp. 241–245, Aug. 2021.

[24] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. Audio Engineering Society (AES)*, (Paris, France), Feb. 2000.

[25] A. Novák, L. Simon, F. Kadlec, and P. Lotton, "Nonlinear system identification using exponential swept-sine signal," *IEEE Trans. on Instrumentation and Measurement*, vol. 59, pp. 2220–2229, Aug. 2010.

[26] European Broadcasting Union, "Sound quality assessment material - recordings for subjective tests: User's handbook for the EBU SQUAM CD," 2008. [Online]. Available: https://tech.ebu.ch/publications/sqamcd.

[27] C. Veaux, J. Yamagishi, and K. MacDonald, "CSTR VCTK corpus: English multi-speaker corpus for CSTR voice cloning toolkit," *University of Edinburgh. The Centre for Speech Technology Research (CSTR)*, 2017.

[28] Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short-time Fourier transform domain," *IEEE Signal Processing Letters*, vol. 14, no. 5, pp. 337–340, 2007.

[29] S. Doclo, W. Kellermann, S. Makino, and S. E. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," *IEEE Signal Processing Magazine*, vol. 32, pp. 18–30, Mar. 2015.

[30] S. Gannot, E. Vincent, S. Markovich-Golan, and A. Ozerov, "A consolidated perspective on multi-microphone speech enhancement and source separation," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 25, pp. 692–730, Apr. 2017.

[31] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, pp. 4–24, Apr. 1988.

[32] N. Gößling and S. Doclo, "Relative transfer function estimation exploiting spatially separated microphones in a diffuse noise field," in *Proc. International Workshop on Acoustic Signal Enhancement (IWAENC)*, (Tokyo, Japan), pp. 146–150, Sep. 2018.

[33] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-Based Noise Power Estimation With Low Complexity and Low Tracking Delay," *IEEE Trans. on Audio, Speech, and Language Processing*, vol. 20, pp. 1383–1393, May 2012.

[34] N. Gößling and S. Doclo, "RTF-steered binaural MVDR beamforming incorporating an external microphone for dynamic acoustic scenarios," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Brighton, UK), pp. 416–420, May 2019.