MULTICHANNEL AUDIO DATABASE IN VARIOUS ACOUSTIC ENVIRONMENTS

Elior Hadad[1], *Florian Heese*[2], *Peter Vary*[2], *and Sharon Gannot*[1]

[1] Faculty of Engineering, Bar-Ilan University, Ramat-Gan, Israel
[2] Institute of Communication Systems and Data Processing (IND)
RWTH Aachen University, Aachen, Germany

{elior.hadad,sharon.gannot}@biu.ac.il {heese,vary}@ind.rwth-aachen.de

ABSTRACT

In this paper we describe a new multichannel room impulse responses database. The impulse responses are measured in a room with configurable reverberation level resulting in three different acoustic scenarios with reverberation times RT_{60} equals to 160 ms, 360 ms and 610 ms. The measurements were carried out in recording sessions of several source positions on a spatial grid (angle range of -90^o to 90^o in 15^o steps with 1 m and 2 m distance from the microphone array). The signals in all sessions were captured by three microphone array configurations. The database is accompanied with software utilities to easily access and manipulate the data. Besides the description of the database we demonstrate its use in spatial source separation task.

Index Terms— Database, room impulse response, microphone arrays, multi-channel.

1 Introduction

Real-life recordings are important to verify and to validate the performance of algorithms in the field of audio signal processing. Common real-life scenarios may be characterized by their reverberant conditions. High level of reverberation can severely degrade speech quality and should be taken into account while designing both single- and multi-microphone speech enhancement algorithms.

Assuming a linear and time-invariant propagation of sound from a fixed source to a receiver, the impulse response (IR) from the sound source to the microphone entirely describes the system. The spatial sound, which bears localization and directivity information, can be synthesized by convolving an anechoic (speech) signal with the IRs. Accordingly, a database of reverberant room IRs is useful for the research community.

There are several available databases. In [1] and [2] binaural room impulse response (BRIR) databases tailored to hearing aid research are presented. A head and torso simulator (HATS) mannikin is utilized to emulate head and torso shadowing effects in the IRs. A database of IRs using both omnidirectional microphone and a B-format microphone was published in [3]. This database includes IRs in three different rooms, each with a static source position and at least 130 different receiver positions. In [4] measurements of IRs of a room with interchangeable panels were published with two different reverberation times. The IRs were recorded by eight microphones at inter-distances of 0.05 m for 4 source microphone distances where the source is positioned in front of the microphone array. These databases are freely available and have been instrumental in testing signal processing algorithms in realistic acoustical scenarios. However, they are somewhat limited with respect to the scope of the scenarios which can be realized (e.g., a limited number of sources direction of arrivals (DOAs) with respect to the microphone array).

The speech & acoustic lab of the Faculty of Engineering at Bar-Ilan University (BIU) (Fig. 1), is a $6\,\mathrm{m} \times 6\,\mathrm{m} \times 2.4\,\mathrm{m}$ room with reverberation time controlled by 60 panels covering the room facets. This allows to record IRs and test speech processing algorithms in various conditions with different reverberation times. In this paper we introduce a database of IRs measured in the lab with eight microphones array for several source-array positions, several microphone inter-distances in three often encountered reverberant times (low, medium and high). In addition, an example application is presented to demonstrate the usability of this database.

The paper is organized as follows. In Sec. 2 the measurement technique is presented. The database is introduced in Sec. 3. Sec. 4 outlines the availability of the database and describes a new signal processing utility package for easy data manipulation. In Sec. 5 we demonstrate the usability of the database by applying a signal separation algorithm to two sources both impinging upon an array from broadside. Finally, conclusions are drawn in Sec. 6.

Fig. 1: Experiment setup in the Speech & Acoustic Lab of the Faculty of Engineering at Bar-Ilan University.

This work was co-funded by the German federal state North Rhine Westphalia (NRW) and the European Union European (Regional Development Fund).

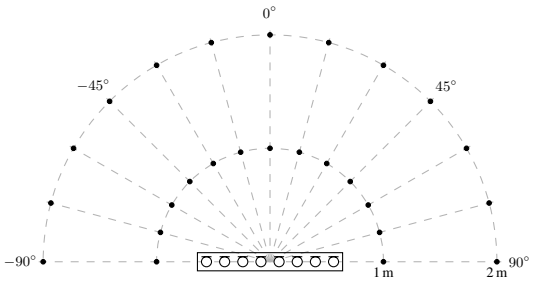


Fig. 2: Geometric setup.

2 Measurement Technique

The measurement equipment consists of RME Hammerfall DSP Digiface sound-card and RME Octamic (for Microphone Pre Amp and digitization (A/D)). The recordings were carried out with an array of 8 microphones of type AKG CK32. As a signal source we used Fostex 6301BX loudspeakers which has a rather flat response in the frequency range 80Hz-13kHz. The software used for the recordings is MATLAB. All measurement were carried out with a sampling frequency of 48 kHz and resolution of 24-bit.

A common method for transfer function identification is to play a deterministic and periodic signal from the loudspeaker $x(t)$ and measure the response $y(t)$ [5]. Due to the input signal periodicity, the input and the output are related by a circular convolution. Accordingly, the IR $h(t)$ can be estimated utilizing the Fourier transform and inverse Fourier transform:

$$h(t) = IFFT\left[\frac{FFT(y(t))}{FFT(x(t))}\right] \quad (1)$$

In [6] it was claimed that in quiet conditions the preferred excitation signal is a sweep signal. The BIU Speech & Acoustics Lab is characterized by such quiet conditions. Moreover, sweeps as excitation signals show significantly higher immunity against distortion and time variance compared to pseudo-noise signals [7]. The periodic excitation signal was set to be a linear sine sweep with a length of 10 s repeated 5 times. The first output period was discarded and the remaining 4 were averaged in order to improve the signal to noise ratio (SNR).

3 Database Description

The measurement campaign consists of IRs characterizing various acoustic environments and geometric constellations. The reverberation time is set (by changing the panel arrangements) to 160 ms (low), 360 ms (medium) and 610 ms (high) to emulate typical acoustic environments, e.g., a small office room, meeting room and a lecture room. An individual geometric microphone spacing and an acoustic condition (reverberation time) defines a single recording session. The loudspeakers are distributed on a spatial grid around the array and are held static for all recording sessions.

The loudspeakers are positioned on two half circles with different radii around the center of the microphone array. The schematic setup is depicted in Fig. 2. To cover a wide range of spatial and acoustic scenarios, the database encompasses nine different recording sessions each of which comprises 26 8-channel impulse responses. In Table 1 detailed measurement conditions are given.

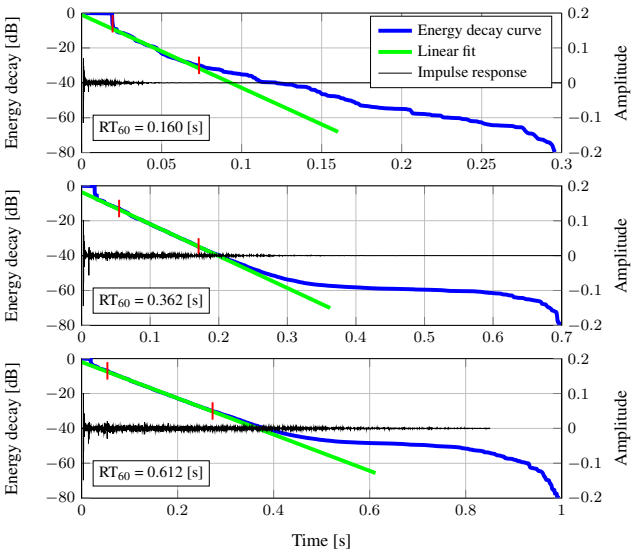


Fig. 3: Energy decay curve for different reverberation times (measured by SP.signal MATLAB class).

For each recording session the acoustic lab was configured by flipping panels and the reverberation time was measured. To ensure a good acoustic excitation of the room, a B&K 4295 omnidirectional loudspeaker was utilized and an estimate of the reverberation time was calculated at five different locations in the room using the WinMLS software [8]. The noise level in silence of the lab was measured as 21.4 dB SPL A-weighted.

An example of measured IRs and their corresponding energy decay curves is depicted in Fig. 3 for three different reverberation times at a distance of 2 m from the source and an angle 0^o. The reverberation times are calculated from the energy decay curves using the Schroeder method [9]. The bounds for the least square fit are marked by red lines.

4 Availability & Tools

All IRs of the database are stored as double-precision binary floating-point MAT-files which can be imported directly to MATLAB. Since the number of IRs is huge, a MATLAB signal processing utility package (SP) was created which allows a simple handling of the database.

The package consists of a signal class (SP.signal) and tools which easily allows to handle multichannel signals and to create

Reverberation time (RT_{60})	160 ms, 360 ms, 610 ms
Microphone spacings	[3, 3 ,3, 8, 3, 3, 3] cm, [4, 4 ,4, 8, 4, 4, 4] cm, [8, 8 ,8, 8, 8, 8, 8] cm
Angles	$-90°$: $90°$ (in $15°$ steps)
Distances (radius)	1m, 2m

Table 1: Measurement campaign properties.

rt60(ch, bound_start, bound_end, plot_it)
Returns RT_{60} reverberation time for channel ch using the Schroeder method [9]. Bound_start and bound_end define the region for the least square fit while plot_it will provide the energy decay curve including the linear fit plot.
to_double
Exports SP.signal to MATLAB matrix.
cut(start_sample, end_sample)
Cuts SP.signal from start_sample to end_sample.
conv
Convolution of two SP.signal (e.g., a clean speech signal and a multichannel impulse response).
resample(new_fs)
Returns a resampled SP.signal with sample rate new_fs.
write_wav(filename)
Exports SP.signal to a .wav-file.

Table 2: Main methods of MATLAB SP.signal class.

spatial acoustic scenarios with several sources by convolution and superposition. The SP.signal class can handle typical entities (speech and audio signals, impulse responses, etc.) and provides several properties such as the sample rate, number of channels and signal length. Supported SP.signal sources are MATLAB matrices and files (.wav and .mat). It is also possible to generate signals like silence, white noise or sinus oscillations using a built-in signal generator. Any additional information like system setup, scenario description or hardware equipment can be stored as metadata. SP.signal also implements the default parameters (plus, minus, times, rdivide, etc.). Further details are listed in Table 2, Table 3 and via MATLAB help command [1].

SP.loadImpulseResponse(db_path, spacing, angle, d, rt60)
Loads an impulse response from db_path folder according to the parameters microphone. spacing, angle, distance and reverberation time and returns the IR as SP.signal.
SP.truncate(varargin)
Truncates each passed SP.signal to the length of the shortest one.
output = SP.adjustSNR(sigA, sigB, SNR_db)
Returns the mixed SP.signal output according to the parameter SNR_dB. It consists of sigA plus scaled version of sigB, where sigA and sigB belong to SP.signal class. For, e.g. evaluation, sigA and the scaled version of sigB are stored in the metadata of output.

Table 3: Tools of MATLAB SP package.

5 Speech Source Separation

In this section we exemplify the utilization of the database. For that, we have considered a scenario with two speech sources, both impinging upon a microphone array from the broadside, with the desired source located behind the interference source. In addition, the environment is contaminated by a directional stationary noise.

[1]The MATLAB tools, sample scripts and the impulse response database can be found at: `http://www.ind.rwth-aachen.de/en/research/tools-downloads/multichannel-impulse-response-database/` and `http://www.eng.biu.ac.il/gannot/`

We apply the subspace-based transfer function linearly constrained minimum variance (TF-LCMV) algorithm [10]. A binaural extension of this algorithm exists [11]. A comparison between the TF-LCMV algorithm and another source separation method utilizing this database can be found in [12].

The M received signals $z_m(n)$ are formulated in a vector notation, in the short-time Fourier transform (STFT) domain as $\boldsymbol{z}(\ell,k) \triangleq \left[\begin{array}{ccc} z_1(\ell,k) & \ldots & z_M(\ell,k) \end{array}\right]^T$ where ℓ is the frame index and k represents the frequency bin. The beamformer output is denoted $y(\ell,k) = \boldsymbol{w}^H(\ell,k)\boldsymbol{z}(\ell,k)$ where the beamformer filters denoted $\boldsymbol{w}(\ell,k) = \left[\begin{array}{c} w_1(\ell,k),\ldots,w_M(\ell,k) \end{array}\right]^T$.

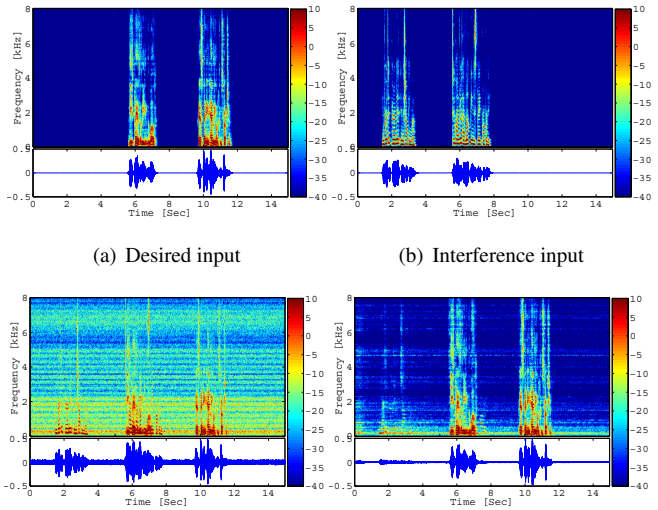
The TF-LCMV is designed to reproduce the desired signal component as received by the reference microphone, to cancel the interference signal component, while minimizing the overall noise power at the beamformer output. It is constructed by estimating separate basis vectors spanning the relative transfer functions (RTFs) of the desired and interference sources. These subspaces are estimated by applying the eigenvalue decomposition (EVD) to the spatial correlation matrix of the received microphone signals. This procedure necessitates the detection of time-segments with nonconcurrent activity of the desired and interference sources. The IR and its respective acoustic transfer function (ATF) in reverberant environment consist of a direct path, early reflections and a late reverberation. An important attribute of the TF-LCMV is its ability to take into account the entire ATFs of the sources including the late reverberation. When two sources impinge upon the array from the same angle, the direct path is similar while the entire ATF differs. Unlike classical beamformers that ignores the reverberation tail, the TF-LCMV takes it into consideration. It is therefore, capable of separating sources that are indistinguishable by classical beamformers.

The test scenario comprises one desired speaker, 2 m from the microphone array, and one interference speaker, 1 m from the microphone array, both at angle 0^o, and one directional stationary pink noise source at angle 60^o, 2 m from the microphone array. The microphone signals are synthesized by convolving the anechoic speech signals with the respective IRs. The signal to interference ratio (SIR) with respect to the non-stationary interference speaker and the SNR with respect to the stationary noise were set to 0 dB and 14 dB, respectively. The sampling frequency was 16kHz. The signals were transformed to the STFT domain with frame length of 4096 samples and 75% overlap. The ATFs relating the sources and the microphone array which are required for the TF-LCMV algorithm can be obtained in one of two ways, i.e., either by utilizing the known IRs form the database or by blindly estimating them from the received noisy recording [10, 11]. The performance in terms of improvement in SIR and improvement in SNR are examined for different scenarios. For evaluating the distortion imposed on the desired source we also calculated the log spectral distortion (LSD) and segmental SNR (SSNR) distortion measures relating the desired source component at the reference microphone, namely $\boldsymbol{e}_1^H\boldsymbol{z}_d(\ell,k)$, and its corresponding component at the output, namely $y_d = \boldsymbol{w}^H(\ell,k)\boldsymbol{z}_d(\ell,k)$, where $\boldsymbol{e}_1$ is M dimensional vector with '1' in the mth component for mth reference microphone and '0' elsewhere, and $\boldsymbol{z}_d(\ell,k)$ denotes the desired source component as received by the microphones. The three reverberation times are tested. We have used the microphone array configuration $[8,8,8,8,8,8,8]$ cm, utilizing either all 8 microphones or only 4 microphones of them (microphones #3-6).

The performance measures are summarized in Table 4. It is evident that the algorithm significantly attenuates the interference speaker as well as the stationary noise for all scenarios. The algorithm's performance for all three reverberation levels is comparable. It is worthwhile explaining these results, as at the first glance, one

Scenario			Performance measures			
T_{60} [s]	ATF	M	ΔSIR	ΔSNR	ΔLSD	ΔSegSNR
$160m$	Real	8	22.15	20.12	0.96	12.41
$160m$	Est	8	20.96	21.32	2.06	9.81
$160m$	Real	4	18.04	16.93	1.34	8.70
$160m$	Est	4	14.73	15.76	2.01	7.51
$360m$	Real	8	22.27	16.77	1.61	8.57
$360m$	Est	8	21.06	19.18	2.62	8.02
$360m$	Real	4	15.63	14.27	2.25	7.21
$360m$	Est	4	14.85	18.48	2.50	8.81
$610m$	Real	8	20.23	15.08	2.34	7.50
$610m$	Est	8	19.00	18.72	3.08	7.74
$610m$	Real	4	14.15	15.40	2.89	4.12
$610m$	Est	4	14.16	17.22	2.74	6.41

Table 4: SNR, SIR improvements, SSNR and LSD in dB relative to microphone reference as obtained by the beamformer for 8 microphone array and 4 microphone array configurations. Three reverberation times are considered. The RTFs required for the beamformer are obtained in one of two ways: either from the true IRs or from the estimated correlation matrices.



(a) Desired input (b) Interference input

(c) Noisy input (d) Enhanced output

Fig. 4: Sonograms and waveforms. The beamformer is utilizing microphones #3-6. The RTFs are extracted from the estimated correlation matrices. RT_{60} equals to 360 ms.

would expect significant performance degradation when reverberation level increases. This degradation does not occur due to the distinct TF-LCMV attribute, taking the entire ATF into account. Under this model both sources, although sharing similar direct path, undergo different reflection patterns and are hence distinguishable by the beamforming algorithm. When the reverberation level becomes even higher (630 ms) the IRs become too long to be adequately modeled with the designated frame length. Hence, a slight performance degradation is expected. In terms of SIR improvement, SNR improvement and SSNR 8 microphone array outperforms 4 microphone array. It can be seen that the LSD measure improves (lower values indicate less distortion) when utilizing the real ATFs instead of estimating them.

Fig. 4 depicts the sonograms and waveforms at various points in the signal flow using 4 microphones, i.e., microphones #3-6. The desired signal, the interference signal and the noisy signal as recorded by microphone #3 are depicted in Fig. 4(a), in Fig. 4(b) and in Fig. 4(c), respectively. The output of the beamformer is depicted in Fig. 4(d). It is evident that the algorithm is able to extract the desired speaker while significantly suppressing the interfering speaker and the noise.

6 Conclusions

We have presented a new multichannel array database of room IRs created in three array configurations. Each recording session consists of 26 sources spatially distributed around the center of the array (1m and 2m distance, angle range of $-90^o : 90^o$ in 15^o resolution). All the sessions where carried out in three reverberation levels corresponding to typical acoustic scenarios (office, meeting and conference room). An accompanying MATLAB utility package to handle the publicly available database is also provided. The usage of the database was demonstrated by a spatial source separation example with two sources impinging upon the array from the broadside.

References

[1] H. Kayser, SD 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 Proc.*, p. 6, 2009.

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

[3] R. Stewart and M. Sandler, "Database of omnidirectional and B-format room impulse responses.," in *IEEE International Conference on Acoustics speech and Signal Processing (ICASSP)*, 2010, pp. 165–168.

[4] J.Y.C. Wen, N.D. Gaubitch, E.A.P. Habets, T. Myatt, and P.A. Naylor, "Evaluation of speech dereverberation algorithms using the MARDY database," in *Proc. Int. Workshop on Acoustic Signal Enhancement (IWAENC)*, 2006.

[5] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *the 108th AES convention*, 2000.

[6] G.B. Stan, J.J. Embrechts, and D. Archambeau, "-Comparison of different impulse response measurement techniques," *Journal of Audio Engineering Society*, vol. 50, no. 4, 2002.

[7] S. Müller and P. Massarani, "Transfer-function measurement with sweeps," *Journal of Audio Engineering Society*, vol. 49, no. 6, pp. 443–471, 2001.

[8] Morset Sound Development, "WinMLS, The measurement tool for audio, acoustics and vibrations," `http://http://www.winmls.com/`, 2004, [Online; accessed 31-March-2014].

[9] M. Schroeder, "New method of measuring reverberation time," *J. of the Acoustical Society of America*, vol. 37, no. 3, pp. 409–412, 1965.

[10] S. Markovich, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant environment with multiple interfering speech signals," *IEEE Trans. Audio, Speech and Language Proc.*, vol. 17, no. 6, pp. 1071–1086, Aug. 2009.

[11] E. Hadad, S. Gannot, and S. Doclo, "Binaural linearly constrained minimum variance beamformer for hearing aid applications," in *Proc. Int. Workshop on Acoustic Signal Enhancement (IWAENC)*, Sep. 2012.

[12] F. Heese, M. Schäfer, P. Vary, E. Hadad, S. Markovich-Golan, and S Gannot, "Comparison of supervised and semi-supervised beamformers using real audio recordings," in *the 27th convention of the Israeli Chapter of IEEE*, Eilat, Israel, Nov. 2012.