

A BINAURAL ROOM IMPULSE RESPONSE DATABASE FOR THE EVALUATION OF DEREVERBERATION ALGORITHMS

Marco Jeub, Magnus Schäfer, and Peter Vary

Institute of Communication Systems and Data Processing (**ind**)
RWTH Aachen University, Germany

{jeub,schaefer,vary}@ind.rwth-aachen.de

ABSTRACT

This paper describes a new database of binaural room impulse responses (BRIR), referred to as the Aachen Impulse Response (AIR) database. The main field of application of this database is the evaluation of speech enhancement algorithms dealing with room reverberation. The measurements with a dummy head took place in a low-reverberant studio booth, an office room, a meeting room and a lecture room. Due to the different dimensions and acoustic properties, it covers a wide range of situations where digital hearing aids or other hands-free devices can be used. Besides the description of the database, a motivation for using binaural instead of monaural measurements is given. Furthermore an example using a coherence-based dereverberation technique is provided to show the advantage of this database for algorithm evaluation. The AIR database is being made available online.

Index Terms— Binaural, Room Impulse Response, Dummy Head, Speech Enhancement, Dereverberation

1. INTRODUCTION

Over the last years the reduction of room reverberation has received increasing attention from the research community. Many authors suggested algorithms for dereverberation which are suitable for hands-free devices, automatic speech recognition systems and digital hearing aids. With increasing computational power and decreasing energy consumption of mobile devices, the integration into commercial devices is forthcoming. Especially for hearing-impaired people, dereverberation can improve the communication in everyday life. The recently introduced wireless-link between the hearing aids on both ears makes the integration of binaural algorithms possible, even though the data rate on this binaural link is still limited.

For the evaluation of dereverberation algorithms, often room impulse responses (RIR) are used which have been generated artificially based on parameters of the acoustic environment or have been measured in real rooms. However, when it comes to the application to digital hearing aids, these RIRs are not suitable because they usually neglect the shadowing effect of the human head.

This paper presents a comprehensive database of measured binaural room impulse responses (BRIR). The recordings took place in a low-reverberant studio booth, an office room, a meeting room and a lecture room, all measured with and without a dummy head. This data is being made available online to the research community¹. The description of binaural speech enhancement systems is out of the scope of this paper. The reader is referred to e.g. [1, Chapter 14]. An

¹The room impulse response database AIR can be found at <http://www.ind.rwth-aachen.de/~bib/jeub09a>

extensive study of signal enhancement including dereverberation algorithms for hearing instruments with normally hearing and hearing impaired listeners has recently been published in [2].

The remainder of this paper is organized as follows. In Section 2 an overview of the common room impulse response generation methods is given. Section 3 discusses the need for binaural room impulse responses. Afterwards, the measurement system will be explained in Section 4 and an evaluation of the AIR database will be presented in Section 5. Finally, conclusions are drawn in Section 6.

2. ROOM IMPULSE RESPONSES

Assuming a linear time-invariant system (LTI), the room impulse response can be used to completely describe the acoustic properties of a room in terms of sound propagation and reflections for a specific source-microphone configuration. Given a set of room impulse responses $h_j(k)$, where $j = 1, \dots, M$, M the total number of microphones and the anechoic speech or audio signal $s(k)$, the reverberant microphone signals can be obtained by

$$x_j(k) = s(k) * h_j(k), \quad (1)$$

where $*$ indicates convolution. In order to obtain the RIR for a specific room, two different methods are mostly used in practice. The first one comprises a simulative approach using the well-known source-image method as proposed by Allen et al. [3]. The main advantage of this method is the many different degrees of freedom. Nearly any “shoebox-like” geometry, source-microphone configuration and microphone type can be simulated.

However, the underlying model has one major limitation. Since the number of generated reflections is limited, the energy decay can differ from the well-known exponential decay model by Polack [4], which is widely used in model-based dereverberation. This limitation can be overcome by the measurement of room impulse responses in real environments. They can be obtained very effectively with pseudorandom sequences (e.g. Maximum length sequences (MLS), see Subsection 4.2 for details, or perfect sequences (PSEQs) cf. [1, Chapter 7]). The main advantage is the accurate reproduction of the acoustic properties of the measurement room. Nevertheless, once the RIR is measured, no changes of the configuration can be made. The use of a reasonably large database can help to lower the impact of this restraint and finally lead to a superior performance of the measured RIRs.

Recently, the so-called Multichannel Acoustic Reverberation Database at York (MARDY) has been presented [5]. It consists of real measured room impulse responses of a room with interchangeable panels. By doing so, the acoustic properties can be changed quite easily. The authors measured the RIR at different source-microphone distances with eight microphones at adjacent distances of 0.05 m. The recordings are well-suited for the evaluation of multichannel dereverberation algorithms. But when it comes to binaural algorithms, this database is not appropriate. One critical aspect is the shadowing effect of the human head which will be described briefly in the next section.

For the sake of completeness, the use of real recordings in reverberant rooms instead of convolving an anechoic signal with the RIR has to be mentioned as an alternative. The main advantage of this procedure is the possibility to completely capture all aspects of the acoustic system omitting the LTI assumption that is necessary for the convolution-based concept described above. However, the assumption of linearity is valid as a first approximation for most real-life scenarios where at least short-term time invariance can be expected as well. The small remaining advantage does definitely not outweigh the immense loss in flexibility during algorithm development and evaluation.

3. BINAURAL ROOM IMPULSE RESPONSES

In the following we will show that the use of binaural room impulse responses is advantageous for the evaluation of dereverberation algorithms for digital hearing instruments. It is still a shortcoming of several publications to use stereo instead of binaural recordings. In the following the term binaural is used for measurements using an artificial head with microphones positioned close to the pinna.

To motivate the use of BRIRs, the effects on the coherence function will be described exemplarily in this section along with the consequences for coherence-based dereverberation. In an ideal diffuse sound field the magnitude squared coherence (MSC) between two closely spaced microphones can be expressed by

$$MSC(f, d) = \left(\frac{\sin(2\pi f d/c)}{2\pi f d/c} \right)^2, \quad (2)$$

where d denotes the distance between two omnidirectional microphones, f the sampling frequency and $c = 340$ m/s the sound velocity [6]. Figure 1 shows the theoretical coherence function for different microphone distances. It can be seen that the distance between the two microphones has great influence on the coherence function. For very close microphones, the function has relatively large values even for higher frequencies. Coherence-based dereverberation algorithms, e.g. [7], rely on a low MSC of the diffuse sound. However, if the microphones have a distance which corresponds to the distance from the left to the right ear of approximately $d = 0.15$ m, the coherence function shows values of < 0.2 for frequencies above 700 Hz. The consequence of Figure 1 is that this class of algorithms requires a certain distance between the two microphones and will always perform poorly for low frequencies. This would exclude this class of algorithms from the application to digital hearing aids with closely spaced microphones per device, and even from devices with a wireless link between the hearing aid of left and right ear. However, it has been shown e.g. in [8, 9] that the latter assumption does not hold if a head is in-between. In the following experiment we show how the coherence changes in the presence of a dummy head. The dotted and dashed lines in Figure 2 represent the measured coherence function of the lecture room ($d_{LM} = 10.2$ m, see Subsection 4.3).

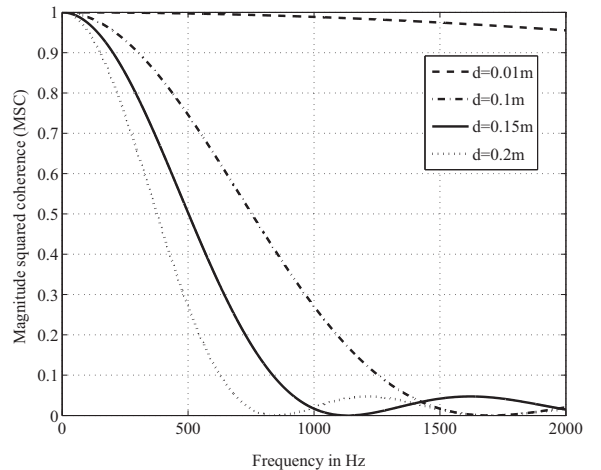


Fig. 1. Theoretical coherence function of two closely spaced microphones for different distances.

The coherence function was estimated from the BRIR using the method described in [10]. It can be seen that the coherence tends to smaller values and hence, algorithms taking advantage of this effect, work more efficiently in lower frequencies. Therefore we conclude

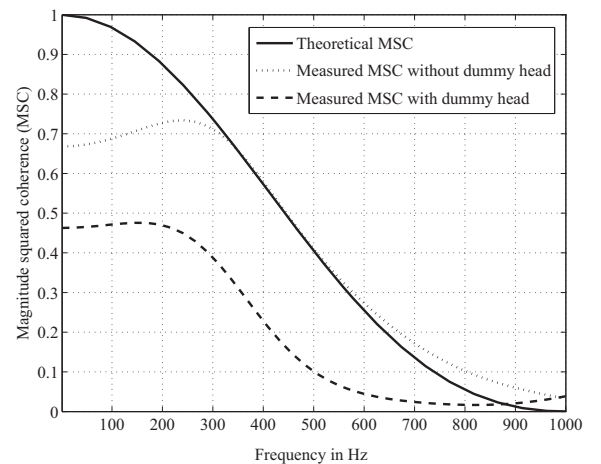


Fig. 2. Influence of the presence of a dummy head on the magnitude squared coherence. Solid line: theoretical MSC using Eq.(2) with $d = 17$ cm. Dashed line: Estimated MSC with a dummy head. Dotted line: Estimated MSC without a dummy head. The two microphones are positioned close to the pinna at a distance of $d = 17$ cm.

that there is a need for binaural room impulse responses. In this paper, we provide a comprehensive database of measured binaural room impulse responses covering several types of situations. An application example on how the performance of the coherence-based Allen dereverberation algorithm [7] improves by means of real binaural room impulse responses is evaluated in Subsection 5.2.

| | Studio booth | Office room | Meeting room | Lecture room |
|-----------------|-----------------------------------|------------------------------|--|---|
| Room dimensions | 3.00 m x 1.80 m x 2.20 m | 5.00 m x 6.40 m x 2.90 m | 8.00 m x 5.00 m x 3.10 m | 10.80 m x 10.90 m x 3.15 m |
| h_L | 1.2 m | 1.2 m | 1.2 m | 1.2 m |
| h_M | 1.2 m | 1.2 m | 1.2 m | 1.2 m |
| d_{MM} | 0.17 m | 0.17 m | 0.17 m | 0.17 m |
| d_{LM} | 0.5 m, 1.0 m, 1.5 m | 1.0 m, 2.0 m, 3.0 m | 1.45 m, 1.7 m, 1.9 m, 2.25 m, 2.8 m | 4.0 m, 5.56 m, 7.1 m, 8.68 m, 10.2 m |
| Wall surface | custom-made low-reflective panels | Glass windows, concrete | Glass windows, concrete | 3x glass windows, 1x concrete wall |
| Floor cover | Carpet | Carpet | Carpet | Parquet |
| Furniture | - | Wooden desk, shelves, chairs | Wooden conference table, bookshelves | Wooden tables, chairs |

Table 1. Properties and measurement configuration of the different rooms.

4. MEASUREMENT SYSTEM

4.1. Measurement Equipment

The measurements were carried out with professional audio equipment in order to retrieve high-quality and low-noise results. The test sequences were emitted by the 2-way active studio monitor Genelec 8130 having a digital AES/EBU interface. For the sound capturing two Beyerdynamic MM1 omnidirectional condenser measurement microphones and the RME Octamic II microphone amplifier were used in combination with an RME Multiface II audio interface. All measurement were performed with a sampling frequency of 48 kHz and 24-bit accuracy. For the binaural recordings the HMS2 artificial head by HEAD acoustics is used. The two microphones are positioned close to the pinna at 1 cm from the ear canal. The measurements were repeated in an otherwise unchanged experimental setup after the artificial head was removed to examine the influence of the head.

4.2. Maximum Length Sequences

As a measurement signal, we used pseudorandom sequences, specifically Maximum Length Sequences which are known to offer some desirable properties. Their crest factor (the ratio between peak power and average power) is at the theoretical optimum of 1, their spectrum is flat (apart from the DC value), their auto correlation has the shape of a Dirac pulse which simplifies the calculation of the impulse response into a circular cross correlation and they are deterministic which allows for summing and averaging of multiple repetitions. Each doubling of the number of repetitions is known to give an additional signal-to-noise gain against uncorrelated noise of 3 dB [6]. For our measurements, we used MLS of degree 15, which gives a measurement length of $2^{15} - 1 = 32767$ samples resulting in a maximum length for the impulse responses of 0.68 s (for the lecture room we used degree 16 leading to a maximum length of 1.36 s). The signals were repeated 17 times and the impulse response was then calculated from the average of the last 16 repetitions resulting in a gain of 12 dB.

4.3. Measuring Rooms

The experiments have been carried out in four room, all with different dimensions and acoustic properties. The following abbreviations are used in table 1. h_L loudspeaker height, h_M microphone

height, d_{MM} microphone-microphone distance, d_{LM} loudspeaker-microphone distance. All room dimensions are given in the form 'length \times width \times height' in meter.

- Low-Reverberant Studio Booth

The most important characteristic of the low-reverberant studio booth is a very low and almost constant reverberation time over frequency and a high acoustical transmission isolation to protect against disturbance noise from outside. This has been achieved by a special arrangement and material of the acoustic panels inside the booth.

- Office room

The second room is a typical office room with standard office furniture.

- Meeting room

This room is a good example to reproduce the acoustic behavior of different speakers during a meeting. Dummy head and measurement system are placed on a fixed position and the room impulse responses are captured for different positions opposite the head. By means of this configuration, the reflection of the conference table, which is very characteristic for these situations, and hence the decrease in speech intelligibility can be examined. The floor plan is depicted in Figure 3(a).

- Lecture room

The largest room which has been measured is a lecture room with typical auditorium furniture like desks and chairs. The loudspeaker was placed at the lectern and the microphone system at different rows with increasing distance to the speaker. This can be seen as a representation of a typical announcement, speech, talk or lecture situation. Refer to Figure 3(b) for details.

4.4. Postprocessing

Since the measurements of the room impulse responses cause different delays (e.g. sound driver, AD/DA converter), a delay compensation is necessary. The provided data is free of such delays. However, the delay due to the sound wave propagation was not removed. The database files are stored as double-precision binary floating-point MAT-files and 24-bit PCM quantized WAV-files.

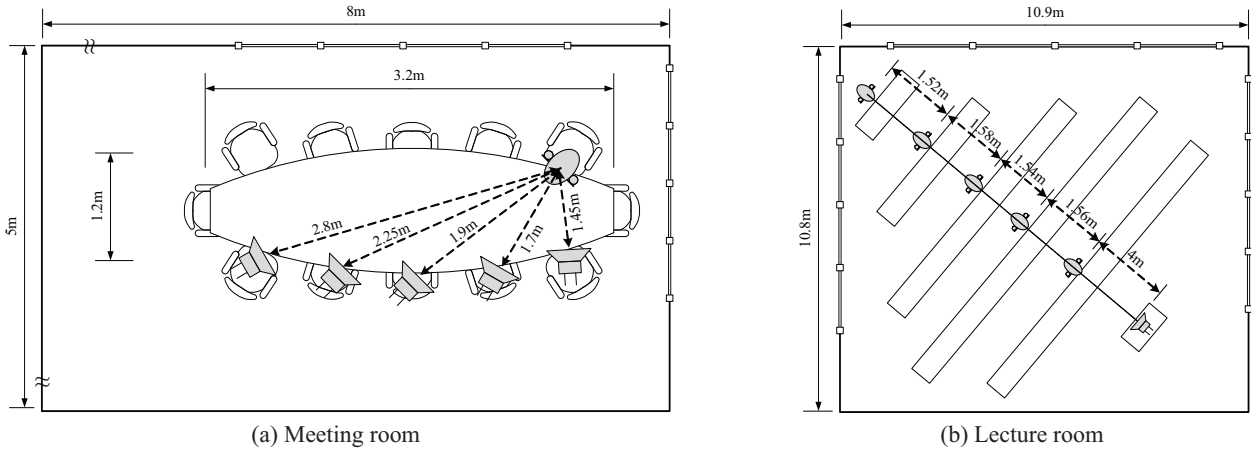


Fig. 3. Room properties and measurement setup for the meeting and lecture room (not to scale).

5. EVALUATION

5.1. Reverberation Times

One fundamental parameter of room acoustics is the reverberation time RT_{60} . It is defined as the time period a switched off sound need to decrease by 60 dB and usually measured with the Schroeder method [11]. In the following the sound wave propagation will be included in the reverberation time. Table 2 shows the reverberation time for each room on every measuring position. A mean value \overline{RT}_{60} for each room is calculated as the average over all positions.

| | d_{LM} in m | RT_{60} in s |
|---|---------------|----------------|
| Studio booth ($\overline{RT}_{60} = 0.12$ s) | 0.50 | 0.08 |
| | 1.00 | 0.11 |
| | 1.50 | 0.18 |
| Office room ($\overline{RT}_{60} = 0.43$ s) | 1.00 | 0.37 |
| | 2.00 | 0.44 |
| | 3.00 | 0.48 |
| Meeting room ($\overline{RT}_{60} = 0.23$ s) | 1.45 | 0.21 |
| | 1.70 | 0.22 |
| | 1.90 | 0.21 |
| | 2.25 | 0.24 |
| | 2.80 | 0.25 |
| Lecture room ($\overline{RT}_{60} = 0.78$ s) | 2.25 | 0.70 |
| | 4.00 | 0.72 |
| | 5.56 | 0.79 |
| | 7.10 | 0.80 |
| | 8.68 | 0.81 |
| | 10.2 | 0.83 |

Table 2. Overview of the measured reverberation times for each room depending on the source-microphone distance. All values are averages between left and right channel. The overall mean values \overline{RT}_{60} are given by the average among all distances.

Since most rooms do not have a constant reverberation time over frequency a reverberation time in subbands is another significant parameter. Figure 4 shows the reverberation time as a function of fre-

quency calculated by means of the Schroeder method and a 1/3-octave filterbank. It can be seen that the reverberation time for the studio booth is almost constant with values of approx. 150 ms over all frequencies. All other rooms show an increasing RT_{60} for decreasing frequencies. The lecture room performs fairly constant for frequencies below 3 kHz. This leads to a first conclusion that this non-constant characteristic must be taken into account when evaluating dereverberation algorithms. Lower frequencies are omitted in the figure due to different reasons. First, the loudspeaker has a cut-off frequency at 55 Hz and show a non-linear frequency response in the range of 55 – 65 Hz. Secondly, in the case of the low-reverberant booth, frequencies below 150 Hz can cause resonance of the system.

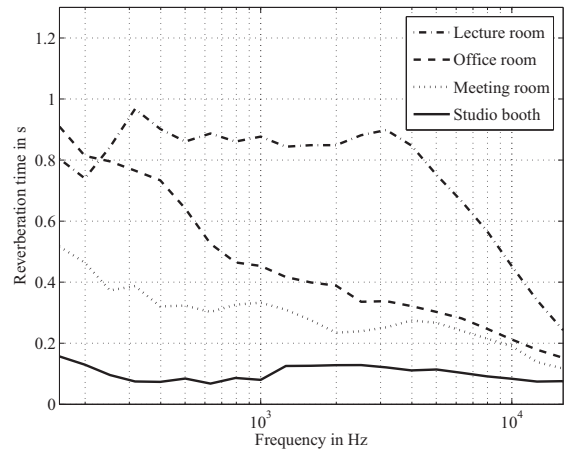


Fig. 4. Subband reverberation time obtained by a 1/3-octave filterbank and the Schroeder method [11]. The values are averaged over all measured positions in the room including averaging of left and right channel.

5.2. Application example

The performance of the Allen dereverberation algorithm [7] using a measured room impulse response (Lecture room, $d_{LM} = 10.2$ m) with and without the dummy head will be compared in the following. This binaural algorithm basically takes advantage of a different coherence for the direct and reflective parts of the captured signals. First, the delay between the coherent and reverberant part is removed and afterwards the signals are added. The entire process is performed in subbands.

For the evaluation a non-intrusive measurement based on the Speech to Reverberation Modulation energy Ratio (SRMR) [12] and the Segmental Signal-to-Reverberation Ratio (SegSRR) is used. Silence periods have been removed before evaluation using the voice activity detector (VAD) of the AMR-WB speech codec [13]. The signal levels are normalized to -26 dBov using the ITU-T Rec. P.56 speech voltmeter [14]. The experiments are carried out on 100 speech files of the TIMIT database sampled at $f_s = 16$ kHz. From Table 3 it can be seen that the dereverberation performs better in the presence of the dummy head. This can be explained with the considerations in Section 3, where it has been shown that the descent of the coherence function of a diffuse sound field becomes steeper due to the shadowing effect of the head.

| | Δ SRMR | Δ SegSRR |
|---------------|---------------|-----------------|
| No dummy head | +1.34 | +0.67 dB |
| Dummy head | +2.16 | +1.39 dB |

Table 3. Reverberation reduction by the dual-channel Allen algorithm with and without the presence of a dummy head in the lecture room. The values indicate the improvement compared to the reverberant speech.

6. CONCLUSIONS

In this paper the Aachen Impulse Response (AIR) database is presented. The main purpose is the evaluation of speech enhancement algorithms and dereverberation algorithms for speech and audio signals in particular. The measurements have been carried out and analyzed in different room. Except for the low-reverberant studio booth, all rooms cause a frequency dependent acoustic behavior and hence the use of subband evaluation criteria is suggested. It has been shown that binaural measurements are essential for the evaluation of binaural dereverberation algorithms due to the shadowing effects of a human head. Exploiting these effects can cause measureable and audible improvements. This statement was strengthened by a case study with the coherence-based dereverberation algorithm by Allen et al. We conclude that binaural room impulse responses give more reliable results when algorithms for digital hearing aids or hands-free devices (e.g. head-sets) are developed and evaluated.

7. REFERENCES

- [1] R. Martin, U. Heute, and C. Antweiler, Eds., *Advances in Digital Speech Transmission*, Wiley&Sons, Chichester, 2008.
- [2] K. Eneman, H. Luts, and J. Wouters, "Evaluation of signal enhancement algorithms for hearing instruments," in *Proc. European Signal Processing Conference (EUSIPCO)*, Lausanne, Switzerland, 2008.
- [3] J.B. Allen and D.A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, No. 4, pp. 943–950, 1979.
- [4] J.-D. Polack, *La transmission de l'énergie sonore dans les salles*, Ph.D. thesis, Université du Maine, Le Mans, France, 1988.
- [5] 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 Echo and Noise Control (IWAENC)*, Paris, France, 2006.
- [6] H. Kuttruff, *Room Acoustics*, Taylor & Francis, London, 2000.
- [7] J.B. Allen, D.A. Berkley, and J. Blauert, "Multimicrophone signal-processing technique to remove room reverberation from speech signals," *J. Acoust. Soc. Am.*, 1977.
- [8] M. Dörbecker, "Sind kohärenzbasierte Störgeräuschreduktionsverfahren für elektronische Hörhilfen geeignet? Modelle zur Beschreibung der Kohärenzeigenschaft," in *Proc. ITG-Fachtagung Sprachkommunikation*, Dresden, Germany, 1998.
- [9] M. Dörbecker, *Mehrkanalige Signalverarbeitung zur Verbesserung akustisch gestörter Sprachsignale am Beispiel elektronischer Hörhilfen*, Ph.D. thesis, RWTH Aachen University, Aachen, Germany, 1998.
- [10] J. Benesty, J. Chen, and Y. Huang, "Estimation of the coherence function with the MVDR approach," in *Proc. IEEE Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Toulouse, France, 2006, vol. 3, pp. III–III.
- [11] M.R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Am.*, vol. 37, no. 3, pp. 409–412, 1965.
- [12] T.H. Falk and W.-Y. Chan, "A non-intrusive quality measure of dereverberated speech," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, USA, 2008.
- [13] TS 26.194, *Adaptive Multi-Rate - Wideband speech codec, Voice Activity Detector*, V6.0.0, 3GPP, 2004.
- [14] ITU-T Rec. P.56, *Objective measurement of active speech level*, ITU, 1993.