The Single- and Multichannel Audio Recordings Database (SMARD)

Nielsen, Jesper Kjær; Jensen, Jesper Rindom; Jensen, Søren Holdt; Christensen, Mads Græsbøll

Published in:
2014 14th International Workshop on Acoustical Signal Enhancement (IWAENC)

DOI (link to publication from Publisher):
10.1109/IWAENC.2014.6953334

Publication date:
2014

Document Version
Accepted author manuscript, peer reviewed version

Link to publication from Aalborg University

Citation for published version (APA):
THE SINGLE- AND MULTICHANNEL AUDIO RECORDINGS DATABASE (SMARD)

Jesper Kjær Nielsen†, Jesper Rindom Jensen‡, Søren Holdt Jensen†, and Mads Gresbøll Christensen‡

†Aalborg University
Dept. of Electronic Systems
{jkn,shj}@es.aau.dk

‡Aalborg University
Audio Analysis Lab, AD:MT
{jrj,mgc}@create.aau.dk

ABSTRACT

A new single- and multichannel audio recordings database (SMARD) is presented in this paper. The database contains recordings from a box-shaped listening room for various loudspeaker and array types. The recordings were made for 48 different configurations of three different loudspeakers and four different microphone arrays. In each configuration, 20 different audio segments were played and recorded ranging from simple artificial sounds to polyphonic music. SMARD can be used for testing algorithms developed for numerous applications, and we give examples of source localisation results.

Index Terms— Multichannel recordings, audio database, source localisation

1. INTRODUCTION

The processing of single- and multichannel audio recordings is a very active field of research within the signal processing community and important in numerous applications such as noise reduction, echo cancellation, source separation, source localisation, room geometry estimation, distributed array processing, recognition, and de-reverberation [1–11]. Despite the heavy research activities in the field, only a few high quality multichannel audio recordings have been made available online. Consequently, many of the developed signal processing algorithms are in research papers only evaluated on simulated data (e.g., generated using room impulse response generators [12, 13]), on data of a low quality, or on non-public data, and this might inhibit reproducibility [14], complicate algorithm comparison, or ultimately lead to the wrong conclusions. There are multiple reasons for why it is difficult to obtain good quality audio recordings. For example, making high quality recordings can be very time-consuming; professional measurement equipment is often expensive; dedicated listening rooms, anechoic chambers, and laboratory equipment might not be available; or many external nuisances, which might be hard to eliminate, can influence the quality of the recordings.

To extend the amount of freely available single- and multichannel audio recordings, we here present a new database called the Single- and Multichannel Audio Recordings Database (SMARD) which is made freely available online. The database contains both single- and multichannel recordings of artificial signals as well as of reverberant and anechoic speech, vocal, and musical signals emitted by various loudspeaker types. The recordings are made with four different microphone arrays in a box-shaped listening room. Moreover, the position of the microphones and loudspeakers inside this room as well as the temperature are also measured. Although useful in many applications, SMARD was compiled for the primary purpose of source localisation and room geometry estimation. For source localisation, SMARD extends databases such as the multi-channel Wall Street journal audio visual corpus (MC-WSJ-AV) [15] and the audio-visual corpus for speaker localization and tracking [16] by using multiple array topologies and other source signals than speech. Moreover, since the microphone and loudspeaker locations relative to the room were also measured, SMARD can also be used for room geometry estimation. Although the data used in [4] have been made available online, SMARD is to the best of our knowledge the first freely available database which can be used for this purpose. For different array structures in different settings, multichannel impulse responses can be found in online available databases. In contrast to these, SMARD contains the raw recordings from which multichannel impulse responses can be estimated if desired.

In this paper, we first describe how the data in SMARD were recorded. Specifically, we describe the listening room, the measurement equipment, and the measurement configurations. Secondly, we give a few examples of source localisation results based on these data.

2. DESCRIPTION OF SMARD

SMARD contains approximately 18 GB of audio recordings at a sampling frequency of 48 kHz. All or a subset of these recordings can be downloaded from the SMARD website.

---

1. The data can be downloaded at http://www.smard.es.aau.dk. If you use SMARD in a peer-reviewed research paper or in a book, we kindly ask you to email us a BibTeX entry which we will publish on the SMARD website.

2. See http://www.commsp.ee.ic.ac.uk/ for a list of these databases.
2.1. The Room

The recordings have all been made in a 60 m$^2$ multichannel listening room at Aalborg University, and a sketch of this room is given in Fig. 1. The room is box-shaped, symmetrical, and has an adjustable reverberation time between approximately 0.2 and 0.4 seconds. To increase the amount of reverberation, we removed the carpet on the floor during the measurement campaign.

2.2. Equipment

Recordings were made for various combinations of different loudspeakers and microphone arrays. As detailed in Table 1, three loudspeakers were used. The Brüel & Kjær OmniPower 4296 and the Brüel & Kjær OmniSource 4295 are both approximately omnidirectional loudspeakers within a limited frequency range. The OmniPower 4296 loudspeaker can emit more sound power than the OmniSource 4295, but can only be considered omnidirectional over a narrower frequency range. The directional loudspeaker is a conventional 3” speaker in a wooden cabinet. Pictures of it and its on-axis impulse response can be found on the SMARD website.

Up to 22 G.R.A.S. microphones were used in various array configurations. The simplest array was just a single microphone which can been seen on the right hand side of Fig. 2a. The other array types were uniform linear arrays (ULA) (see Fig. 2c), uniform circular arrays (UCAs) (see Fig. 2d), and an orthogonal array (see Fig. 2b). Finally, a dummy microphone, which is simply a capacitor mounted on the microphone pre-amplifier, was also present in all recordings. The recordings from the dummy microphone can be used to inspect electrical noise, cross-talk, etc. Except for the loudspeakers, the microphones, and the arrays, all measurement equipment was situated in a control room adjacent to the multichannel listening room. A list of this equipment can be found in Table 1.

2.3. Measurements Configurations

Measurements were made for a total of 48 different configurations. Each configuration is enumerated by a four digit number of the form ABCD where the first and most significant digit A denotes the type of loudspeaker; the second digit B denotes the position and orientation of the loudspeaker; the

---

**Table 1**: Measurement Equipment. More details can be found on the SMARD website.

<table>
<thead>
<tr>
<th>Loudspeakers</th>
<th>Microphones</th>
<th>Arrays</th>
<th>A/D Converters</th>
<th>Sound Card</th>
<th>Power amplifier</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brüel &amp; Kjær OmniPower 4296</td>
<td>G.R.A.S. Prepolarized Free Field Microphone 40AZ</td>
<td>Single microphone</td>
<td>Behringer Ultragain Pro-8 Digital ADA8000</td>
<td>RME Digi 9652 Project Hammerfall</td>
<td>Rotel RB-976</td>
<td>Playrec 2.1.0</td>
</tr>
<tr>
<td>Brüel &amp; Kjær OmniSource 4295</td>
<td>G.R.A.S. 26CC Microphone pre-amplifier</td>
<td>Three 7-element ULA with 5 cm spacing</td>
<td>Portaudio v. 19</td>
<td>Steinberg ASIO SDK 2.3</td>
<td>Portaudio v. 19</td>
<td>MATLAB 2013b</td>
</tr>
<tr>
<td>Custom-made 3” directional loudspeaker</td>
<td>(dummy) Bruel &amp; Kjær JJ-2617 Coaxial input adapter (51 pF)</td>
<td>Two 6-element UCAs with radii of 4 cm and 6 cm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2**: The 48 measurement configurations.

<table>
<thead>
<tr>
<th>Artificial Sounds</th>
<th>Speech/vocal signals</th>
<th>Musical signals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Five seconds of silence</td>
<td>8 Soprano vocal from the EBU SQAM CD</td>
<td>14 Clarinet from the EBU SQAM CD</td>
</tr>
<tr>
<td>2 Exponential sine sweep from 10 Hz to 24 kHz</td>
<td>9 Quartet vocal from the EBU SQAM CD</td>
<td>15 Trumpet from the EBU SQAM CD</td>
</tr>
<tr>
<td>3 Harmonic signals with increasing fundamental frequency in steps</td>
<td>10 Male voice from the EBU SQAM CD</td>
<td>16 Xylophone from the EBU SQAM CD</td>
</tr>
<tr>
<td>4 Eight repetitions of a 16th order MLS sequence</td>
<td>11 Child’s voice from the TSP speech database</td>
<td>17 Abba excerpt from the EBU SQAM CD</td>
</tr>
<tr>
<td>5 Pink noise</td>
<td>12 Female voice from the TSP speech database</td>
<td>18 Bass flute from the MIS database</td>
</tr>
<tr>
<td>6 Single sinusoidal tone with increasing frequency in steps</td>
<td>13 Male voice from the TSP speech database</td>
<td>19 Guitar from the MIS database</td>
</tr>
<tr>
<td>7 White Gaussian noise</td>
<td></td>
<td>20 Violin from the MIS database</td>
</tr>
</tbody>
</table>

**Table 3**: The 20 audio segments.

(ULA) (see Fig. 2c), uniform circular arrays (UCAs) (see Fig. 2d), and an orthogonal array (see Fig. 2b). Finally, a dummy microphone, which is simply a capacitor mounted on the microphone pre-amplifier, was also present in all recordings. The recordings from the dummy microphone can be used to inspect electrical noise, cross-talk, etc. Except for the loudspeakers, the microphones, and the arrays, all measurement equipment was situated in a control room adjacent to the multichannel listening room. A list of this equipment can be found in Table 1.

---

**Fig. 1**: A sketch of the multi-channel listening room.
third digit C denotes the type(s) of microphone arrays; and the least significant digit D denotes the position and orientations of these arrays. Table 2 summarises these configurations and further details can be found on the SMARD website. As an example, Fig. 2a shows a picture of configuration number 0002 which include the OmniPower 4296, the orthogonal array, the single microphone, and the dummy microphone.

2.4. Audio Segments

For each of the 48 configurations, a total of 20 audio segments were played and recorded. As listed in Table 3, seven artificial signals, six speech/vocal signals, and seven musical signals were used. The artificial signals were all created in MATLAB and the code for generating them can be found at the SMARD website. The speech and musical signals consist of both reverberant and anechoic signals. The signals from the EBU SQAM CD [17] are reverberant signals whereas the signals from the TSP speech database [18] and the musical instrument samples (MIS) database [19] are anechoic signals. All of these databases are freely available online for research usage.

For every configuration, the temperature inside the multi-channel listening room was measured and stored before these 20 audio segments were played. For each of the audio segments, all of the microphone recordings and a loop-back of the loudspeaker signal were stored in the database. A pause of two seconds was added between the segments to ensure that the sound field within the room was stationary before the next segment was played. The first audio segment was just five seconds of silence. The recordings made with this input signal can be used to inspect the stationary acoustic background noise.

3. EXAMPLES OF USE

As previously mentioned, SMARD is useful for evaluating, e.g., noise reduction, localisation, and room geometry estimation algorithms. In this section, we present some results obtained from the evaluation of two localisation algorithms on some of the data in SMARD. The evaluated algorithms are the steered response power with phase transform (SRP-PHAT) method [20], and a near-field, maximum likelihood (ML) method recently proposed in [21]. As the ML method assumes that the desired signal is quasi-periodic, the methods were applied on the synthetic harmonic signals and the violin signals. More specifically, a single segment of 100 samples
was used from each microphone in different configurations. The segment from the harmonic signals was taken from the last part of the signal where the pitch is 500 Hz, while the segment from the violin signals was taken from the first part. The pitch of the signals, which is needed in the ML method, is estimated using the method in [22], and the number of harmonics was assumed known. Further details about the simulation setup can be found on the SMARD website along with the code for running the simulations.

With this simulation setup, we first of all obtained the results in Figure 3, depicting cost functions and location estimates for the SRP-PHAT and ML methods when applied for localisation of the violin source in configuration 2000. To obtain these plots of the cost functions versus two coordinates at a time, the last coordinate was fixed to the value estimated by the method. We see from the results that the cost functions peak relatively close to the true source position. The methods were also evaluated on other scenarios, yielding the results in Table 4. Generally, the angle (azimuth $\varphi$ and elevation $\psi$) estimates are close to the true angles except for a few cases where a strong reflection from the wooden floor dominates the cost functions (configurations 2001 and 2003). The range ($r_c$) estimates are more inaccurate, but in most cases, the source is also placed relatively far from the arrays. These results support the potential of applying SMARD for evaluation of, e.g., localisation methods, and the validity of the recorded data.

### 4. CONCLUSIONS

We have here presented the Single- and Multichannel Audio Recordings Database (SMARD) which can be used for testing algorithms developed for numerous audio signal processing tasks such as source localisation and room geometry estimation. SMARD is made freely available online to facilitate easier testing on real recordings, reproducibility of results, and algorithm comparison based on the same data. The database contains multichannel recordings for 20 audio segments in 48 different configurations arising from using three different loudspeakers, four different microphone arrays, and various source and sensor locations inside a box-shaped listening room.

### Acknowledgements

The authors would like to thank Claus Vestergaard Skipper, Martin Bo Møller, Neo Kaplanis, and Peter Skotte for assisting in solving practical problems w.r.t. making the measurements.

---

**Table 4:** Location estimates in spherical coordinates for different configurations.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>True $\varphi$</td>
<td>111.6</td>
<td>111.1</td>
<td>130.2</td>
<td>139.4</td>
<td>108.4</td>
<td>134.8</td>
<td>111.6</td>
<td>111.1</td>
<td>130.2</td>
<td>139.4</td>
<td>108.4</td>
<td>134.8</td>
</tr>
<tr>
<td>True $\psi$</td>
<td>3.1</td>
<td>-2.3</td>
<td>-3.2</td>
<td>2.8</td>
<td>-3.9</td>
<td>2.9</td>
<td>3.1</td>
<td>-2.3</td>
<td>-3.2</td>
<td>2.8</td>
<td>-3.9</td>
<td>2.9</td>
</tr>
<tr>
<td>True $r_c$</td>
<td>5.3</td>
<td>5.3</td>
<td>3.8</td>
<td>5.9</td>
<td>4.6</td>
<td>4.8</td>
<td>5.3</td>
<td>5.3</td>
<td>3.8</td>
<td>5.9</td>
<td>4.6</td>
<td>4.8</td>
</tr>
<tr>
<td>SRP-P $\varphi$</td>
<td>112.7</td>
<td>111.8</td>
<td>131.3</td>
<td>141.1</td>
<td>99.1</td>
<td>130.8</td>
<td>112.7</td>
<td>111.1</td>
<td>131.4</td>
<td>140.4</td>
<td>112.5</td>
<td>147.1</td>
</tr>
<tr>
<td>SRP-P $\psi$</td>
<td>-0.7</td>
<td>-27.2</td>
<td>-4.4</td>
<td>-22.1</td>
<td>-15.1</td>
<td>15.0</td>
<td>2.6</td>
<td>-1.7</td>
<td>-7.6</td>
<td>-0.4</td>
<td>3.7</td>
<td>4.2</td>
</tr>
<tr>
<td>SRP-P $r_c$</td>
<td>4.4</td>
<td>2.8</td>
<td>5.6</td>
<td>2.4</td>
<td>4.3</td>
<td>2.9</td>
<td>6.7</td>
<td>6.7</td>
<td>6.0</td>
<td>7.9</td>
<td>2.8</td>
<td>3.9</td>
</tr>
<tr>
<td>ML-AP $\varphi$</td>
<td>111.7</td>
<td>112.1</td>
<td>131.2</td>
<td>140.3</td>
<td>112.8</td>
<td>133.7</td>
<td>112.4</td>
<td>110.9</td>
<td>131.4</td>
<td>139.9</td>
<td>105.8</td>
<td>138.1</td>
</tr>
<tr>
<td>ML-AP $\psi$</td>
<td>-0.4</td>
<td>-26.1</td>
<td>-4.8</td>
<td>-20.3</td>
<td>-12.2</td>
<td>14.0</td>
<td>4.3</td>
<td>-1.5</td>
<td>-5.2</td>
<td>0.5</td>
<td>1.7</td>
<td>2.5</td>
</tr>
<tr>
<td>ML-AP $r_c$</td>
<td>7.0</td>
<td>3.2</td>
<td>5.2</td>
<td>2.6</td>
<td>6.0</td>
<td>1.6</td>
<td>5.6</td>
<td>3.9</td>
<td>3.0</td>
<td>6.1</td>
<td>8.3</td>
<td>6.1</td>
</tr>
</tbody>
</table>

**Fig. 3:** Cost-functions versus cartesian coordinates for the SRP-PHAT and ML methods when applied on configuration 2000.
5. REFERENCES


