THE SINGLE- AND MULTICHANNEL AUDIO RECORDINGS DATABASE (SMARD)

Jesper Kjær Nielsen†, Jesper Rindom Jensen‡, Søren Holdt Jensen†, and Mads Græsbøll 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 [1]. 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-A V) [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 can also be used for room geometry estimation. For various array structures and settings, estimated multichannel impulse responses can be found in online databases [2]. In contrast to these, SMARD also contains the raw recordings for a large number of source signals.

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

---

1. This work was partially funded by the Danish Council for Independent Research - Grant no.: DFF–1337-00084, the Villum Foundation, and InnovationsFonden.

2. 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 [1]. 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-A V) [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 can also be used for room geometry estimation. For various array structures and settings, estimated multichannel impulse responses can be found in online databases [2]. In contrast to these, SMARD also contains the raw recordings for a large number of source signals.

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.
2.1. The Room

The recordings have all been made in a 60 m² 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 been measured using the Brüel & Kjær Type 2270 to have a reverberation time of approximately 0.15 seconds.

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 be 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 a 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, microphones, and arrays, all measurement equipment was situated in a control room adjacent to the multichannel listening room. The equipment is listed 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
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 a description of how they were generated can be found on 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. Reverberant signals were included in the set of audio segments since the localisation of loudspeakers often involves such source 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 acoustical background noise. From the audio recordings containing the ESS, we have also estimated the room impulse responses (RIR) from the speaker to the different microphones. These, and further details about how they were estimated, are available from the download section of the website.

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 $\phi$ 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.

### 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 $\phi$</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>4.6</td>
<td>5.9</td>
<td>4.8</td>
</tr>
<tr>
<td>SRP-P $\phi$</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 $\phi$</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>3.0</td>
<td>6.1</td>
<td>8.3</td>
<td>6.1</td>
<td>0.4</td>
</tr>
</tbody>
</table>

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

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.
5. REFERENCES


