DEMAND: a collection of multi-channel recordings of acoustic noise in diverse environments

Version 1.0

Joachim Thiemann, Nobutaka Ito, Emmanuel Vincent

June 9, 2013

1 Introduction

Microphone arrays, a (typically regular) arrangement of several microphones, allow for a number of interesting signal processing techniques. The correlation of audio signals from microphones that are located in close proximity with each other can, for example, be used to determine the spatial location of sound source relative to the array, or to isolate or enhance a signal based on the direction from which the sound reaches the array.

Typically, experiments with microphone arrays that consider acoustic background noise use controlled environments or simulated environments. Such artificial setups will in general be sparse in terms of noise sources. Other pre-existing real-world noise databases (e.g. the AURORA-2 corpus [1], the CHiME background noise data [2], or the NOISEX-92 database [3]) tend to provide only a very limited variety of environments and are limited to at most 2 channels.

The DEMAND (Diverse Environments Multichannel Acoustic Noise Database) aims to provide a set of recordings that allow testing of algorithms using real-world noise in a variety of settings. All recordings are made with a 16-channel array, with the smallest distance between microphones being 5 cm and the largest being 21.8 cm.

2 Descriptions of environments

The current database is divided into 6 categories, 4 of which are "inside" and 2 of which are open air. The inside environments are classified as Domestic, Office, Public, and Transportation; the open air environments are Street and Nature. There are 3 environment recordings in each category.

- In the "Domestic" category, the three recordings are:
  - **DWASHING** inside a washroom with a front-loading washer running a wash cycle,
  - **DKITCHEN** inside a kitchen during the preparation of food,
  - **DLIVING** inside a living room.
- The "Nature" category has three outdoor environment recordings:
  - **NFIELD** a sports field with activity nearby,
  - **NRIVER** a creek of running water,
  - **NPARK** a well-visited city park.
- In the "Office" category, the recordings are:
  - **OOFFICE** a small office with a three people using computers,
  - **OHALLWAY** a hallway inside an office building, with individuals and groups passing by occasionally,
  - **OMEETING** a meeting room while the microphone array is discussed.
- In the category "Public", recordings are of interior public spaces:
  - **PSTATION** the main transfer area of a busy subway station,
  - **PCAFETER** a busy office cafeteria,
  - **PRESTO** a university restaurant at lunchtime.
- The "Street" category has outdoor recordings made near inner-city public roads, in the latter two cases predominantly pedestrian traffic:
  - **STRAFFIC** a busy traffic intersection,
  - **SPSQUARE** a public town square with many tourists,
  - **SCAFE** the terrace of a cafe at a public square.

- "Transportation" describes a set of recordings made in the inside of vehicles:
  - **TMETRO** a subway,
  - **TBUS** a public transit bus,
  - **TCAR** a private passenger vehicle.

All recordings were made in Rennes (France) and its immediate vicinity, in the period between May 2012 and February 2013.

Recordings were made at 48 kHz for a long duration then trimmed to 300 s to avoid setup noises that are not part of the "natural" background noise. The data is available both at the original sampling rate (without modification of the samples, merely changing the container format) and downsampled to 16 kHz. For further format descriptions, see below.

3 Microphone array and recording equipment

3.1 Array construction

The Microphone array consists of 16 microphones arranged in 4 staggered rows, spaced such that there is a 5 cm distance from each microphone to its immediate neighbors. The array is in a plane which in all recordings is parallel to the ground.

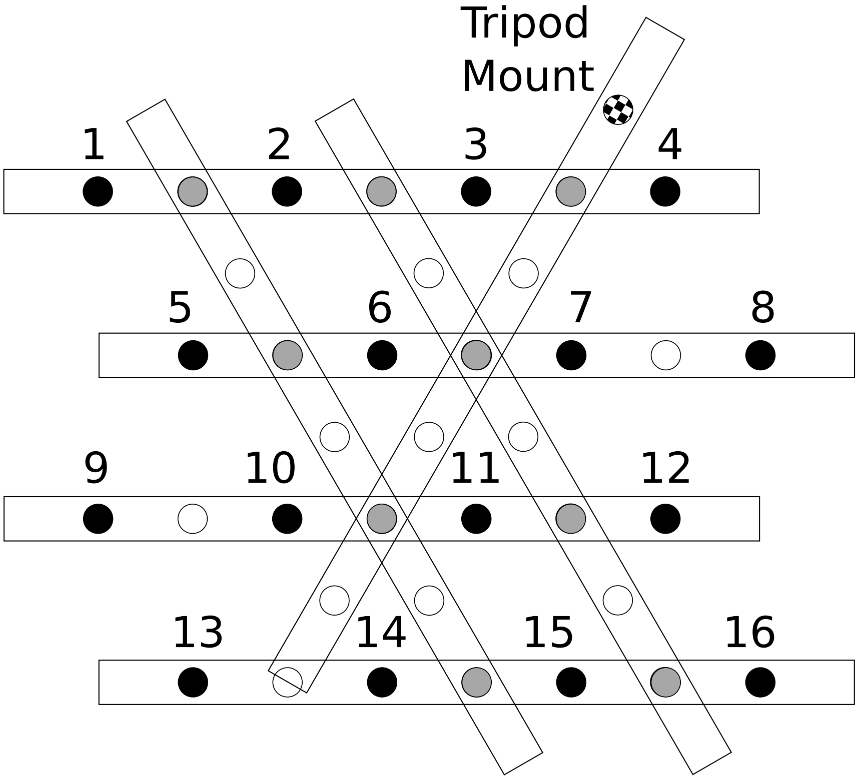

Figure 1: Schematic view of array construction. Black filed circles indicate microphones, grey filled circles show the location of bolts for connecting the bars. Empty circles show unused holes in the bars. The tripod mount is indicated in the top right. The numbers show the channel number of each microphone.

Figure 3.1 shows the design of the array. The array was constructed using 7 U-shaped aluminum bars, 20 cm long and 1.2 cm wide. Holes were drilled at 2.5 cm intervals, used either for mounting microphones or connecting bolts. The MATLAB script `arraypos.m` provided on the website creates a 16×2 MATLAB array with the xy coordinates (in meters) of each channel microphone relative to the microphone of channel 1. Post construction, the microphones were found to be +/- 2 mm of the designed locations.

The array was mounted on a standard microphone tripod, typically suspending the microphones 1.4 m off the ground. The major exception is the TCAR recording, due to space constraints; the array was held at approximately chest height on the passenger seat of the vehicle.

3.2 Recording equipment

The array uses 16 Sony ECM-C10 omnidirectional electret condenser microphones. They are connected to a Inrevium / Tokyo Electron Device TD-BD-16ADUSB USB D/A converter. The converter was connected to laptops running either Microsoft Windows or the Linux operating system; the choice of operating system should not have affected the recordings.

The data was captured using the tools supplied with the USB converter and stored in its custom format[1]. The MATLAB script `ich2wav.m` provided on the website was then used to trim the data and convert it to the standard RIFF (".wav") format.

Each environment noise recording is available as a set of 16 individual mono sound files in a subdirectory (e.g. `DKITCHEN/ch01.wav`), packaged in a "zip" file. The resampling from 48 kHz to 16 kHz was done using the standard "resample()" function in MATLAB R2012a.

4 Additional notes on the data

4.1 DC offset

The DC offset of each channel can be found by a simple average of the PCM data within each channel. We found that the DC offset in all channels is less than the A/D converter step size.

4.2 Gain variations

The recorded signals were not subject to any gain normalization. Therefore, the original noise power in each environment is preserved.

Given the size of the microphone array compared to the distances of the noise sources in each environment, we expect that the overall level of sound at each microphone should be roughly the same, barring occlusion effects from the support structure. However, the microphones of the array are electret microphones, and contain internal preamplifiers. They are not calibrated with respect to each other, and so gain variations are to be expected: we found that the energy in some channels is consistently higher than in other channels. Algorithms working on this data should compensate for this variation.

5 License

6 Acknowledgment

This work was supported by INRIA under the Associate Team Program VERSAMUS (`http://versamus.inria.fr`).

References

[1] `http://catalog.elra.info/product_info.php?products_id=693`

[1] The command to record was `sinichapp -0 0` *name*`.ich` *duration* `-g 0x50`. The first option selects the sampling rate of 48 kHz, the last option sets the gain of the microphone preamps to +20 dB.

[2] http://spandh.dcs.shef.ac.uk/projects/chime/PCC/datasets.html

[3] http://www.speech.cs.cmu.edu/comp.speech/Section1/Data/noisex.html