The Diverse Environments Multi-channel Acoustic Noise Database (DEMAND): A database of multichannel environmental noise recordings

Joachim Thiemann, Nobutaka Ito, Emmanuel Vincent

BACKGROUND

In audio recordings outside of controlled studio setups, the presence of acoustic background noise is a simple fact of life. As a result, there is continued interest in developing methods to disentangle some sound of interest from that background noise [1, 2, 3].

Typically, the development and evaluation of algorithms that separate, reduce, or remove acoustic background noise uses setups with controlled or simulated environments [4]. Such artificial setups are, in general, sparse in terms of noise sources. This is a poor substitute for real acoustic background noise and does not sound "natural" to casual listeners evaluating the performance of the processing techniques being developed.

The solution then is to record noise in various environments, based on the targeted use of the algorithm in question. If these recordings are made available to the research community, such recordings can be used as reference points for researchers to compare each others work. There are now several real-world noise databases, for example the AURORA-2 corpus [5], the CHiME background noise data [6], and the NOISEX-92 database [7]. Unfortunately, these databases provide only a very limited variety of environments, are limited to at most 2 channels and, with the exception of CHiME, are not free.

In current research projects, we are investigating the use of source separation algorithms and beamforming techniques for signal enhancement and acoustic noise suppression/removal where the signal is captured using a multi-microphone array. Since the above mentioned databases do not provide more than two channels, we decided to create our own set of recordings and make these available under a Creative Commons Attribution-ShareAlike 3.0 Unported [8] license for general distribution. These recordings can be found at `http://www.irisa.fr/metiss/DEMAND/`.

PHYSICAL CHARACTERISTICS OF THE MICROPHONE ARRAY

For the recordings, we built a planar array of 16 microphones supported by a structure of metal rods with cross-braces to avoid deformation. The 16 microphones were arranged in 4 staggered rows, with 5 cm spacing of each microphone from its immediate neighbours. Using this arrangement, the array could also be regarded as smaller linear arrays (in three directions) and smaller crystal arrays [9]. In all recordings, the plane of the array was parallel to the ground. The array was mounted on a standard microphone tripod at a height of 1.5 m. The tolerances of the construction were such that the actual locations of the microphones were within 2 mm of the design. Figure 1 presents the schematic of the physical design and shows a photograph of the actual array.

MICROPHONES, AMPLIFIER, AND A/D CONVERTER

The array used 16 Sony ECM-C10 omnidirectional electret condenser microphones. They were connected to an Inrevium / Tokyo Electron Device TD-BD-16ADUSB USB soundcard, which internally used Asahi Kasei AK4563A 16-bit A/D converters with internal preamps. The soundcard was connected to laptops running either Microsoft Windows or the Linux operating system. The choice of operating system did not affect the recordings.

DATABASE DESIGN

The database of recordings was designed to consist of six broad categories, with three environments being recorded within each category. Four of these categories consisted of enclosed

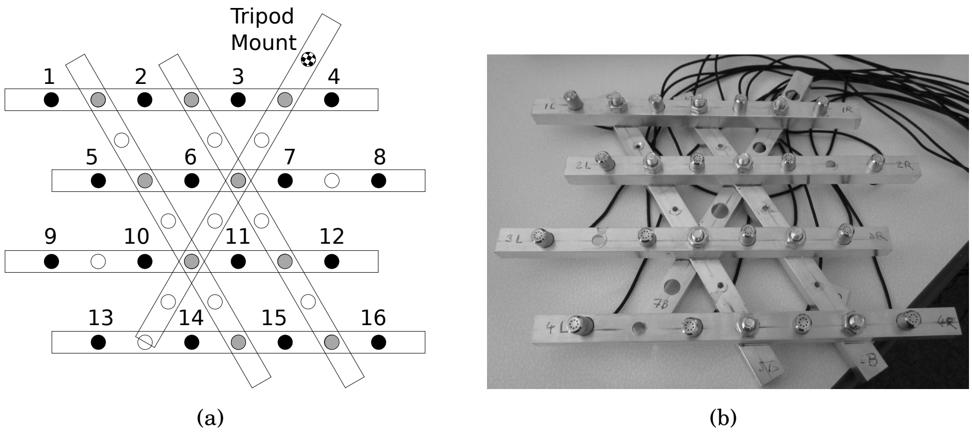

FIGURE 1: (a) Diagram of microphone array layout. Black dots indicate microphone locations with channel numbers. Grey dots show connection bolts. (b) Photograph of microphone array.

spaces, with the remaining two containing recordings done outdoors. The indoor environments were classified as Domestic, Office, Public, and Transportation; the open air environments were Nature and Street. This emphasis on indoor environments reflected the fact that the microphone array used for the recordings was not well suited to outdoor use, especially in less than ideal weather conditions (the array had no protection against rain or wind). Descriptions of the categories and the recordings within each category are given in Table 1.

TABLE 1: Noise Database structure: Categories and recordings in each category

Category	Environment	Description
Domestic	DKITCHEN	inside a kitchen during the preparation of food
	DLIVINGR	inside a living room
	DWASHING	domestic washroom with washing machine running
Office	OHALLWAY	a hallway inside an office building with occasional traffic
	OMEETING	a meeting room while the microphone array is discussed
	OOFFICE	a small office with a three people using computers
Public	PCAFETER	a busy office cafeteria
	PRESTO	a university restaurant at lunchtime
	PSTATION	the main transfer area of a busy subway station
Transportation	TBUS	a public transit bus
	TCAR	a private passenger vehicle
	TMETRO	a subway
Nature	NFIELD	a sports field with activity nearby
	NPARK	a well-visited city park
	NRIVER	a creek of running water
Street	SCAFE	the terrace of a cafe at a public square
	SPSQUARE	a public town square with many tourists
	STRAFFIC	a busy traffic intersection

All recordings were made in Rennes (France) and its immediate vicinity, in the period between May and August 2012. Three environments were omitted from the initial release of the database: DLIVINGR, NRIVER, and SCAFE. These will be added to the database when conditions are suitable to perform more field recordings.

PROPERTIES OF THE SOUND RECORDINGS

The recordings were captured at a sampling rate of 48 kHz and with a target length of 5 minutes (300 s). Actual audio capture time was somewhat longer thereby allowing us to remove set-up noises and other artefacts by trimming. However, the recordings were not spliced yielding a single uninterrupted, contiguous time segment for each.

The recorded signals were not subject to any gain normalization. Therefore, the original noise power in each environment was preserved. Given the size of the microphone array compared to the distances of the noise sources in each environment, we expected that the overall level of sound at each microphone would be roughly equal, barring occlusion effects from the support structure. However, the microphones of the array were electret microphones and contained internal preamplifiers. They were not calibrated with respect to each other, and so gain variations were expected (the data sheet for the microphones specifies a tolerance in sensitivity of 3.5 dB [10]). Table 2 shows the calibration data obtained by placing a 01dB Cal21 calibrator at each microphone and recording the peak amplitude relative to full scale with a 1 kHz 94 dBSPL sinusoidal signal present at each microphone.

TABLE 2: Calibration data for all channels of the array using a 1 kHz 94dB SPL signal at each microphone. Values are given in dB (peak) relative to the full scale of the sample data type.

Channel	1	2	3	4	5	6	7	8
dBFS	-23.4	-23.4	-22.9	-24.2	-22.5	-23.2	-23.6	-23.0
Channel	9	10	11	12	13	14	15	16
dBFS	-22.4	-25.4	-24.5	-22.9	-23.7	-23.0	-23.1	-23.3

Figure 2 shows the loudness profiles for each of the recordings in dBSPL (A-weighted) at microphone one. The loudness was calculated using the "slow" response (window size of about 1 s). The database contains recordings that are very even in terms of loudness (PRESTO, PSTATION, NFIELD) and others that vary by almost 30 dB (DKITCHEN, TMETRO, STRAFFIC). Note that in one recording (TBUS), the signal was slightly clipped in some channels but, since this was mostly due to low-frequency vibration from the ground, it did not show up in the loudness profile.

CONCLUSION

Freely-available noise recordings are a valuable resource to research audio processing algorithms. The DEMAND database is a free Creative Commons licensed set of 16-channel recordings of noise, which to the best of our knowledge is not available from other sources. We hope that this database will prove useful to students and experienced researchers.

ACKNOWLEDGMENTS

This work was supported by INRIA under the Associate Team Program VERSAMUS (`http://versamus.inria.fr`). The authors would also like to thank Dan Freed of EarLens Corp. for advice regarding the calibration.

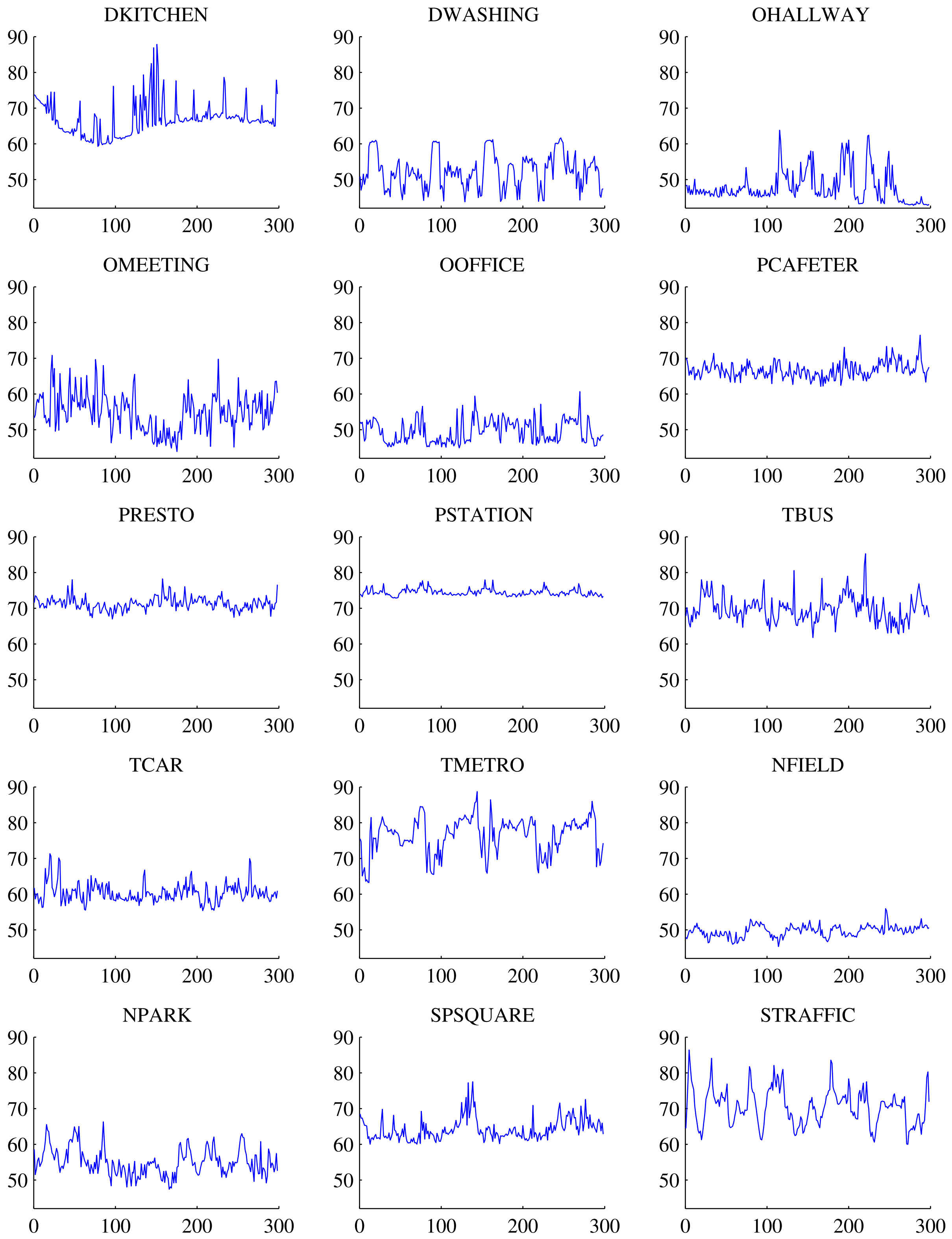

FIGURE 2: Loudness profiles of recordings, in dBSPL (A-weighted), measured at channel one. The horizontal axis shows the time in seconds.

References

[1] M. S. Brandstein and D. B. Ward, eds., *Microphone Arrays: Signal Processing Techniques and Applications* (Springer) (2001).

[2] S. Makino, T.-W. Lee, and H. Sawada, eds., *Blind Speech Separation* (Springer) (2007).

[3] E. Vincent and Y. Deville, "Audio applications", in *Handbook of Blind Source Separation, Independent Component Analysis and Applications*, edited by P. Comon and C. Jutten, 779–819 (Academic Press) (2010), URL `http://hal.inria.fr/inria-00544027`.

[4] E. Vincent, S. Araki, F. J. Theis, G. Nolte, P. Bofill, H. Sawada, A. Ozerov, B. V. Gowreesunker, D. Lutter, and N. Q. K. Duong, "The Signal Separation Evaluation Campaign (2007-2010): Achievements and Remaining Challenges", Signal Processing **92**, 1928–1936 (2012), URL `http://hal.inria.fr/inria-00630985`.

[5] European Language Resources Association, "Aurora project database 2.0", (2008), URL `http://catalog.elra.info/product_info.php?products_id=693`.

[6] J. Barker, E. Vincent, N. Ma, H. Christensen, and P. Green, "The PASCAL CHiME speech separation and recognition challenge", Computer Speech and Language **In Press** (2012), URL `http://spandh.dcs.shef.ac.uk/projects/chime/PCC/datasets.html`.

[7] "Noisex-92 database", (1992), URL `http://www.speech.cs.cmu.edu/comp.speech/Section1/Data/noisex.html`.

[8] Creative Commons, "Creative commons attribution-sharealike 3.0 unported", (2013), URL `http://creativecommons.org/licenses/by-sa/3.0/deed.en_CA`.

[9] N. Ito, H. Shimizu, N. Ono, and S. Sagayama, "Diffuse noise suppression using crystal-shaped microphone arrays", Audio, Speech, and Language Processing, IEEE Transactions on **19**, 2101 –2110 (2011).

[10] Sony Corporation, *ECM-C115/C10/CS10 Electret Condenser Microphone* (2004).