A Novel Aliasing vs Non Aliasing Audio Dataset for Always-On IoT Microphone Experimentation

Jack Adiletta
Worcester Polytechnic Institute
Worcester, MA, USA

Khan Mohammad Nur Hossain
Worcester Polytechnic Institute
Worcester, MA, USA

Matthew Reynolds
Columbia University
New York, NY, USA

Shiwei Fang
Augusta University
Augusta, GA, USA

Bashima Islam
Worcester Polytechnic Institute
Worcester, MA, USA

Abstract

We present the first ever dataset of deliberately aliasing audio. Aliasing, an acoustic phenomenon which adds noise and other abnormalities to a recording, occurs when the sample rate of audio is below the Nyquist sampling rate. The Nyquist sampling rate is defined as 2x the highest frequency present in the audio. Sampling below the Nyquist rate folds the upper frequencies into the lower frequencies causing distortion because the lower sample rate cannot accurately capture the high frequencies. Using three standard audio datasets, we curated several sub-datasets of aliasing audio at different sample rates. This dataset is necessary in the age of IoT because always-on microphone devices will benefit from lower sample rates which require less power to sample the audio and less bandwidth to transmit recordings. Lower sample rates will cause aliasing and our dataset is the first dataset pre-made to test the effects of aliasing on downstream tasks like audio classification. The dataset can also be used to build algorithms to classify audio as aliasing or not.

CCS Concepts

• **Computer systems organization** → **Sensor networks**; Embedded software; • **Computing methodologies** → *Classification and regression trees*; *Neural networks*.

Keywords

adaptive sampling, aliasing audio, signal processing, smart infrastructure

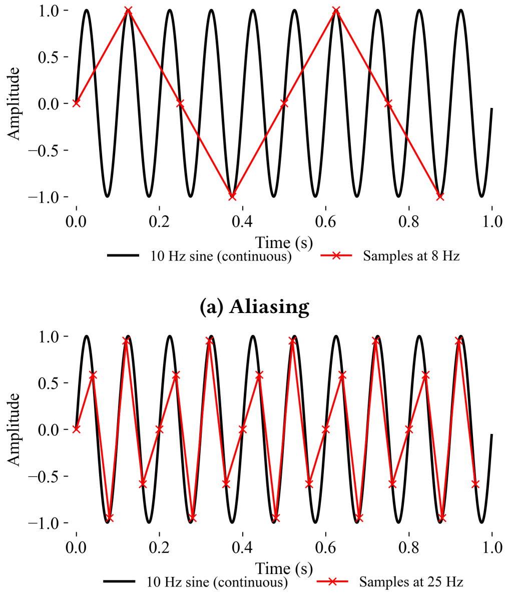
ACM Reference Format:
Jack Adiletta, Khan Mohammad Nur Hossain, Matthew Reynolds, Shiwei Fang, and Bashima Islam. 2025. A Novel Aliasing vs Non Aliasing Audio Dataset for Always-On IoT Microphone Experimentation. In *The Eighth International BuildSys Workshop on DataFM: Data Acquisition & Analysis with Foundational Models (BUILDSYS '25), November 19–21, 2025, Golden, CO, USA*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3736425.3771962

BUILDSYS '25, Golden, CO, USA

ACM ISBN 979-8-4007-1945-5/25/11
https://doi.org/10.1145/3736425.3771962

1 Introduction

Audio aliasing occurs when audio is recorded or played back at a frequency less than *twice* the highest frequency present, called the *Nyquist Frequency* [3, 6]. Audio with frequency elements above 5 kHz when sampled at 10kHz will have 'sonic elements' that distort the recording. Figure 1(a) graphically illustrates when a 10 Hz sine wave is sampled at 8 Hz, which is lower than the Nyquist frequency (20Hz) of the 10 Hz sine wave.



(a) Aliasing

(b) Non-Aliasing

Figure 1: Effect of aliasing on a sinusoid signal.

Figure 1(b) shows a sine wave sampled at 25Hz, above the Nyquist Frequency for a 10Hz Sine wave. The playback of the 25Hz sampled wave would not have aliasing, while the playback of the 8Hz wave would have 'sonic elements.' It is important to note that the 10Hz wave could be lost if sampled at exactly 20Hz if the first sample is taken at 0ms. Without a slight time shift, every sample of the 10Hz wave would be at 0. Thus, signals must be sampled at *greater* than twice the Nyquist Frequency [6].

2 Generating Aliasing and Non Aliasing Audio

We construct a novel dataset using 97,891 unique 1-second clips from UrbanSound8K [8], ESC-50 [7], and MAVD [9], totaling over 27 hours of diverse, real-world audio. From each clip we generated an aliasing and non aliasing version of the clip. The dataset includes aliased and non-aliased versions of each clip across five target sampling rate: 22 kHz, 16 kHz, 8 kHz, 4 kHz, 2 kHz, and 1 kHz, providing broad coverage of real-world environmental acoustic variability[1]. This dataset is used in our proposed method – *EfficientMic* [1] – for adaptive acoustic sensing with a single microphone.

To generate aliasing and non aliasing audio, we developed a novel methodology. First we read the audio digitally at the rate the file was stored at (44 kHz or 22 kHz). We chunk the audio into 1 second clips. We determine if the clip is too quiet to be analyzed (i.e. portions of audio are quiet between dog barks in the dog barking clips) and throw out files with a mean square power less than a configurable significance threshold. We also do not include files with a ratio of mean square power of the frequencies above 11 kHz to the mean square power of all frequencies less than the significance threshold. This check ensures that the files have "strong aliasing" as the mean square power of the high frequencies in the audio make up a significant portion of the total mean square power. Using both checks ensures all files used in the dataset are loud and will alias when under sampled.

We used a 6-pole Butterworth filter from the Scipy python library with a 3dB frequency at half of the five target frequencies to filter each file. The 5 3dB frequencies are 11 kHz, 8 kHz, 4 kHz, 2 kHz, 1 kHz, and 500 Hz. The Nyquist sampling rate of the each sub file filtered at half the target frequency is now the target frequency; thus the filtered files will not have aliasing audio. The filtered files are still sampled at a rate of 22 kHz or 44 kHz, so we use the Scipy library to down sample the filtered files to their respective target sample rate (2 kHz filtered files are sampled at 4 kHz). We also down sample the unfiltered version of the file which will have aliasing as the frequencies above half the sample rate were not removed and will result in noise and distortion.

3 Encoding the Data

Once we have 10 versions of each 1-second clip (clips sampled at 5 target frequencies that are each filtered or unfiltered) from the original three datasets, we save each version as a Short-Time-Fourier-Transform (STFT) and Mel-frequency Cepstral Coefficients (MFCC). Before saving as a final version, we perform a full pass over the entire dataset collecting the minimum and maximum values globally for all the STFTs and the mean and standard deviation for the MFCCs. This way we can min-max normalize the STFTs and Z-score the MFCCs. We do not convert the STFTs to decibels thus the STFTs are non-negative energy values with a wide distribution of values which could harm the performance of an AI/ML model. A min-max normalization of the STFT maintains the positive nature of STFTs while scaling the data to within an acceptable range for AI/ML models. MFCCs contain negative values with value ranges that are specific to each group of MFCC coefficients (one row in the MFCC), thus Z-scoring row wise is an acceptable procedure.

The outcome of these pre-processing steps is shown in figure 2 and 3. As an example we show STFT and MFCC results from an air conditioner, a constant drone with a distribution of frequencies from 1-10 kHz. The top row in both figures is audio sampled at 22 kHz while the bottom row is sampled at 4 kHz. The left column in both figures is the unfiltered, aliasing audio while the right column is the filtered, non aliasing audio. These figures are stored as a numpy array and can be converted back to a 2D representation for processing by an LSTM or other sequential model.

At high sample rates (top row in figures 2,3), the visual representations of the aliasing and non aliasing audio are nearly identical as there are few frequencies above the filter 3 dB frequency. However, when the sample rate decreases (in the bottom row of visualizations), minor differences appear. In particular for the MFCCs, the row corresponding to coefficient group 4 is noticeably lighter in the filtered column than the unfiltered column. Note that color scaling is held constant for the MFCCs and STFTs.

4 Challenges and Limitations

Our methodology for generating the sub datasets of aliasing and non aliasing files at different sample rates is concrete, but we have a major difficulty that we attempted to address that could be expanded in future work: quantifying the strength of the aliasing audio. The dataset is intended to have half the files as aliasing and half as non aliasing; we ensure that only samples which have at least some aliasing (as defined by a significance threshold) are included in both the filtered and unfiltered file collections. However, some file classes like gunshots were much easier to discern if they were aliasing because they contained higher frequencies. Our dataset characterizes files in a binary fashion - aliasing or not - as opposed to the strength of the aliasing. Some files with minimal aliasing may not be great examples of aliasing relative to files with higher frequencies. Our solution to this problem was tuning the significance threshold to eliminate roughly 5% of files which did not have a high enough mean square power of high frequencies (above 11 kHz) relative to the total mean square power. While our heuristic is logical and yielded a successful dataset for our task, future work in the field could yield a stronger metric for selecting files to include based on the strength of the aliasing. As always on devices become more prevalent, experimenting with lower sample rates will be critical to reduce network traffic and power consumption; an aliasing vs non aliasing dataset will be very helpful to assess the efficiency of models with lower sample rates. Our dataset fills the gap in this area.

5 Using the Dataset

The dataset is pre-organized into train-validation-test splits in 80%, 10%, 10% proportions respectively. Each file is named <chunk_index>-<class>.npy, where <class> is the original label extracted from the filename (10 classes for Urbansound 8K, 50 classes for ESC 50). The MAVD city-sound recording corpus does not have a class as the original dataset was not class-based. Numpy file formatting was chosen for its easy integration into Python - Scikit learn machine learning pipelines. The directory structure of the dataset is shown below:

[1] https://zenodo.org/records/16712803

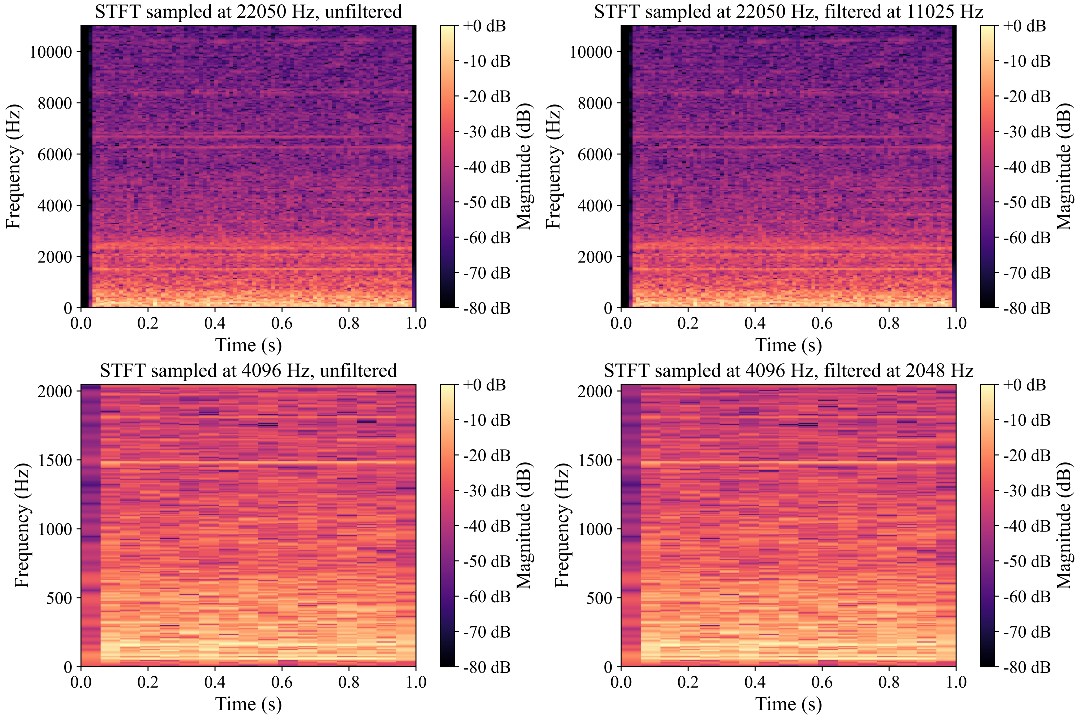


Figure 2: Aliasing Demonstration STFTs with Air Conditioner Audio Example

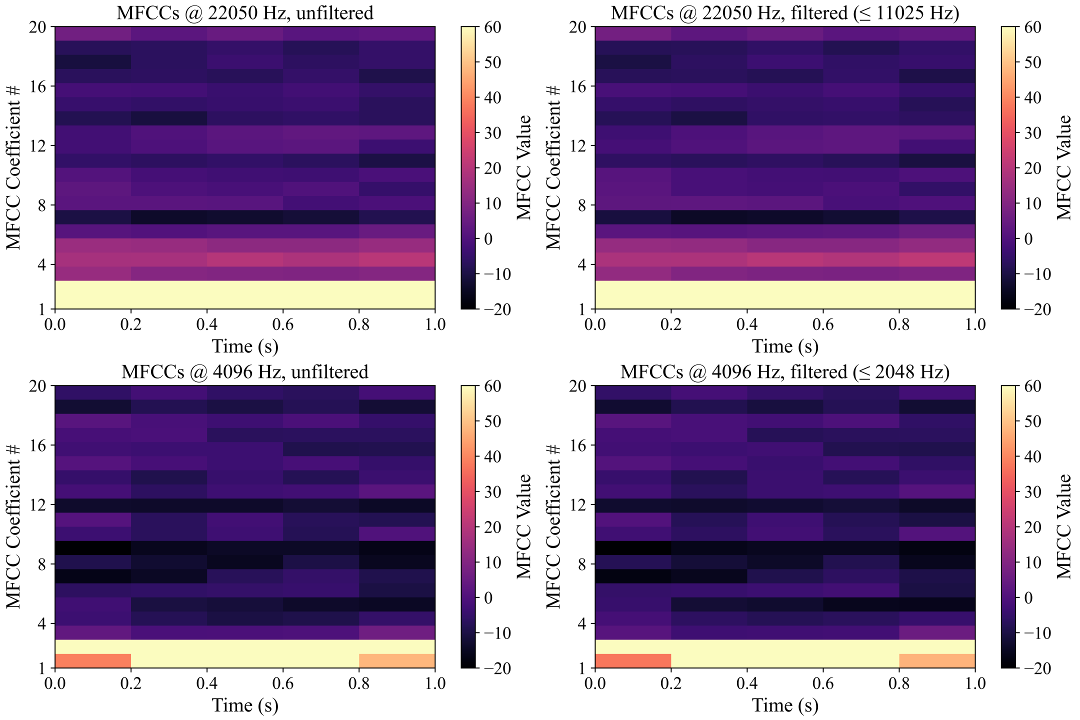


Figure 3: Aliasing Demonstration MFCCs with Air Conditioner Audio Example

```
Processed_Files/
     DS_U8K/ # urbansound 8K
     DS_ESC/ # ESC 50
     DS_ZEN/ # MAVD Dataset
          22000/
          16000/
          8000/
          4000/
          2000/
          1000/
             train/
                 filtered/
                 unfiltered/
             validation/
                 filtered/
                 unfiltered/
             test/
                 filtered/
                 unfiltered/
```

6 Similar Datasets

Our dataset focuses on analyzing audio at different sample rates and the effects of aliasing during that analysis, however, there are other datasets aimed at different problems in the area of always on listening microphones for urban settings. SONYC-UST-V2 provides a variety of urban sound recordings with additional metadata including location and timestamp data [2]. This dataset includes 18510 10 second recordings annotated with 23 tags. The dataset was produced in conjunction with the New York City Department of Environmental Protection (DEP) and represents many of the frequent causes of noise complaints in New York City [2]. CFAD and similar speech datasets provide examples of fake and real speech patterns [5]. CFAD is a Chinese Fake Audio Detection dataset generated by adding noise and generating fake audio with twelve mainstream speech-generation techniques [5]. In [4], the authors present an urban sound corpus which contains both 'events' and 'background' sounds encompassing a wide variety of well documented audio samples. While these datasets allow for comprehensive audio analysis, no dataset focuses on the effects of different, sub-Nyquist sampling rates or allow for this analysis out-of-the-box.

References

[1] Jack Adiletta, Khan Mohammad Nur Hossain, Matthew Reynolds, Shiwei Fang, and Bashima Islam. 2025. EfficientMic: Adaptive Acoustic Sensing with a Single Microphone for Smart Infrastructure. In *Proceedings of the 12th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation.*

[2] Mark Cartwright, Jason Cramer, Ana Elisa Mendez Mendez, Yu Wang, Ho-Hsiang Wu, Vincent Lostanlen, Magdalena Fuentes, Graham Dove, Charlie Mydlarz, Justin Salamon, Oded Nov, and Juan Pablo Bello. 2020. SONYC-UST-V2: An Urban Sound Tagging Dataset with Spatiotemporal Context. arXiv:2009.05188 [cs.SD] https://arxiv.org/abs/2009.05188

[3] Fabian Esqueda, Stefan Bilbao, and Vesa Valimaki. 2016. Aliasing reduction in clipped signals. *IEEE Transactions on Signal Processing* 64, 20 (2016), 5255–5267. doi:10.1109/tsp.2016.2585091

[4] Jean-Remy Gloaguen, Arnaud Can, Mathieu Lagrange, and Jean-Francois Petiot. 2018. Creation of a corpus of realistic urban sound scenes with controlled acoustic properties. *Proceedings of Meetings on Acoustics* 30, 1 (01 2018), 055009. arXiv:https://pubs.aip.org/asa/poma/article-pdf/doi/10.1121/2.0000664/18176538/pma.v30.i1.055009$_1.online.pdf$ doi:10.1121/2.0000664

[5] Haoxin Ma, Jiangyan Yi, Chenglong Wang, Xinrui Yan, Jianhua Tao, Tao Wang, Shiming Wang, and Ruibo Fu. 2023. CFAD: A Chinese Dataset for Fake Audio Detection. arXiv:2207.12308 [cs.SD] https://arxiv.org/abs/2207.12308

[6] Bogdan Mihai. 2009. Sampling rate and aliasing on a virtual laboratory. *Journal of Electrical and Electronics Engineering* 2 (10 2009).

[7] Karol J. Piczak. 2015. ESC: Dataset for Environmental Sound Classification. In *Proceedings of the 23rd Annual ACM Conference on Multimedia* (Brisbane, Australia, 2015-10-13). ACM Press, 1015–1018. doi:10.1145/2733373.2806390

[8] Justin Salamon, Christopher Jacoby, and Juan Pablo Bello. 2014. UrbanSound8K.

[9] Pablo Zinemanas, Pablo Cancela, and Martín Rocamora. 2019. MAVD: A Dataset for Sound Event Detection in Urban Environments. In *Proceedings of the DCASE 2019 Workshop*. New York, USA, 25–26.