JOURNAL OF
EMERGING INVESTIGATORS

Article

Comparison of spectral subtraction noise reduction algorithms

Darshan Shah[1], Bhavesh Shah[2]

[1] Sahyadri School KFI, Pune, Maharashtra, India

[2] Associate Professor of Paediatrics, Government Medical College, Miraj, Maharashtra, India

SUMMARY

Noise in media is any undesirable signal that masks relevant information content. The addition of noise to real-world data in any context is practically inevitable. Noise reduction algorithms in the past have addressed the problem but lacked adaptability to various real-world applications while also being time and resource extensive. Spectral subtraction provides a hybrid approach to noise reduction that incorporates versatility and efficient resource usage. This research tested the performance of two spectral subtraction noise reduction algorithms (stationary and non-stationary) across five categories of real-world noise (speech only, speech with natural noise, music, animal sounds, and noise only). The research question under study was how stationary and non-stationary spectral subtraction algorithms differ in their noise reduction performance when subjected to the various categories of noise. The testing was done based on normalized cross-correlation, which is the similarity between the noise-reduced audio and the original recording in each case. Non-stationary spectral subtraction performed better in samples where human speech was the target: speech only and speech with natural noise. Stationary spectral subtraction performed better when denoising music and animal sounds. This anomaly in performance between the two algorithms was only noted in categories with no human speech. These results exemplify the performance and versatility of different spectral subtraction algorithms. The category-specific results can be used to employ specific spectral subtraction algorithms for specific tasks for optimum performance.

INTRODUCTION

Noise in media is any random, unpredictable, and undesirable signal or change in signals, that masks relevant information content (1). Pertaining to digital audio, noise is any undesirable signal that hinders the quality and intelligibility of the relevant sound signals (1).

The addition of noise to real-world data in any context is practically inevitable and can be traced to a multitude of internal and external factors, some as fundamental as the electronic equipment involved in the system itself. To a lay person, noise may merely be an inconvenience in audio consumption. However, in many fields, noiseless sound is a fundamental necessity. For instance, noise reduction increases the control producers have over the sound they record and produce. Furthermore, in situations involving disaster response, manufacturing and education, the impact of noise can induce delays and errors, affecting real outcomes in public safety, production, and performance (2). For these reasons, noise reduction is a quintessential tool.

Industry-standard filter methods like Butterworth filtering and Wiener Filtering are prevalent methods for noise reduction (3, 4). More modern Artificial Intelligence (AI)-based Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are also in development and use (5). However, traditional filters like the Butterworth and Wiener are not adequately versatile. Versatility refers to the ability of a filtering technique to adapt to a wide range of signal characteristics and noise profiles encountered in different applications. AI-intensive methods, on the other hand, can be time and resource intensive (4, 6).

As is evident by the definition of noise, it is subjective to what the relevant information content is. In different contexts, humans can intuitively define and identify what may or may not be relevant. For instance, in an audio call, anything other than human speech is considered noise. However, in a program built to identify a song, anything other than music, including human speech, is considered noise. Therefore, the definition of noise is dynamic and contextual. To a computer, the task of identifying noise as separate from sound in a compound audio sample is a complex challenge.

The algorithms used in this study were two spectral subtraction noise reduction algorithms: stationary and non-stationary. These algorithms define noise using estimated noise thresholds, or gates, for different frequency channels. This prediction assumes the desired sound signals are uncorrelated with the noise signals and dominant. It then refers to these dominant signals to calculate the noise gates. The noise reduction algorithms used in this study are variations of spectral gating or spectral subtraction algorithms. They rely on the additive, and consequently subtractive, property of sound signals and essentially subtract the noise from the compound signal. This requires the identification of the 'noise' signal, which is done independently in the different frequency bands.

Average signal and noise spectra are estimated in these separate parts of the signal and subtracted from each other so that the average signal-to-noise ratio (SNR) is improved.

Both algorithms tested follow the same algorithm structure up until this step. In stationary spectral subtraction, noise estimates are calculated on each frequency channel to determine a noise gate. Then the gate is applied to the signal (7). Non-stationary spectral subtraction is an extension of the stationary noise reduction algorithm. It follows a similar process but also incorporates the dynamic changing of the

noise gates over time. This capitalizes on the assumption that audio patterns persisting over longer timescales (relative to the timescale of the signal) are noise (7).

Due to its hybrid nature, spectral subtraction incorporates both versatility and efficient resource usage, judiciously employing artificial intelligence as well as traditional gating methods.

We studied the effectiveness of spectral subtraction algorithms in this complex task of identifying and reducing auditory noise. More specifically, the aim was to determine whether non-stationary spectral subtraction is more effective at noise reduction compared to stationary spectral subtraction in five categories of noise:

1. Speech only - human speech with synthetic noise
2. Speech with natural noise - human speech with natural noise
3. Music
    a. Music with speech - music along with human speech
    b. Music without speech - music by itself
4. Animal sounds - non-human vocal sounds
5. The noise only - no distinct target

The findings showed that non-stationary spectral subtraction was more effective in noise reduction when the target was human speech, including speech only, speech with natural noise, and music with speech. On the other hand, stationary spectral subtraction showed slightly better performance in reducing noise in music-only samples and significantly better performance in reducing noise in animal sounds.

RESULTS

The testing was done by normalized cross-correlation, which the similarity between the noise-reduced audio and the original recording in each case.

A Mann-Whitney U test was conducted on the resulting data since it contains outliers and has a relatively small sample space.

The performance of stationary and non-stationary spectral subtraction algorithms was compared across different noise categories using the Mann-Whitney U test. The test results for the speech only, speech with natural noise, music, and animal sounds categories, respectively are presented **(Table 1)**. Notably, the noise-only category does not have a test result due to normalized cross-correlation values of 0 in one or more samples.

The differences in performance were not statistically significant. Despite this, the results are qualitatively different. Based on the normalised cross-correlation, the non-stationary noise reduction algorithm performed better with categories targeting human speech as the subject, such as the speech only, speech with natural noise, and music with speech categories.

Contrary to expectation, the stationary noise reduction algorithm outperformed the non-stationary noise reduction algorithms in the remainder of the categories, including the music without speech, and animal sounds.

The performance of the two algorithms comparatively over the five noise categories is visualized **(Figure 1)**. In the 'speech only', 'speech with natural noise' and 'music with speech' categories, non-stationary spectral subtraction performed better on average. In 'music without speech', and 'animal sounds', stationary spectral subtraction performed better, notably in the latter. In the 'noise only' category, stationary spectral subtraction failed to reduce noise and scored 0%.

The performance of the two algorithms comparatively over all 15 samples is also visualized **(Figure 2)**.

In 9 samples out of 15, non-stationary spectral subtraction performed better than stationary spectral subtraction. The accuracy scores for samples 13 and 14 were 0% for at least one of the two algorithms.

Noise Category	Algorithm	Sample Size	Mean (%)	Standard Deviation (%)	Mann-Whitney U	Z-score	*p*
Speech Only	Stationary	5	80.37	12.69	10	-0.52	0.69
	Non-stationary	5	80.37	9.61			
Speech with natural noise	Stationary	3	74.96	6.99	2	-1.09	0.4
	Non-stationary	3	79.4	3.66			
Music	Stationary	4	65.13	18.81	7	-0.29	0.886
	Non-stationary	4	69.8	7.85			
Animal sounds	Stationary	1	88.68	-	0	-1	1
	Non-stationary	1	34.9	-			

Table 1: Summary of Mann-Whitney U Analysis between noise reduction performance of stationary spectral subtraction and non-stationary spectral subtraction for different noise categories.

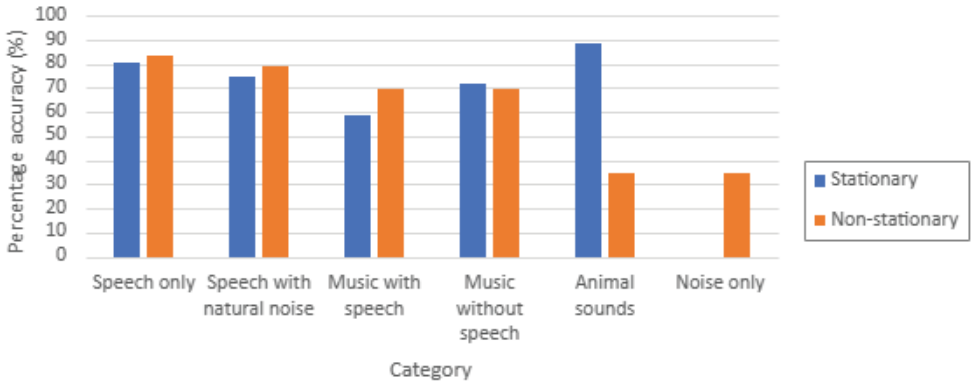

Figure 1: Graphical representation of comparative category wise normalised cross-correlation between S_O and S_R. Across the categories of 'speech only,' 'speech with natural noise,' and 'music with speech,' non-stationary spectral subtraction exhibited superior performance on average. For 'music without speech' and 'animal sounds,' stationary spectral subtraction demonstrated better results, particularly in the case of the latter. Conversely, in the 'noise only' category, stationary spectral subtraction was ineffective in noise reduction, receiving a score of 0%.

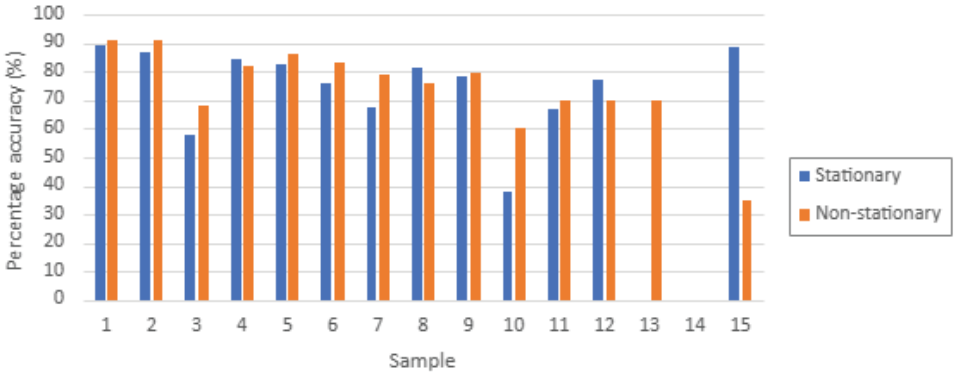

Figure 2: Graphical representation of comparative category wise normalised cross-correlation between S_O and S_R for all samples. Non-stationary spectral subtraction outperformed stationary spectral subtraction in 9 out of 15 instances. For samples 13 and 14, at least one of the two algorithms yielded a 0% accuracy score. The exceptional results observed in samples 13 and 14, as well as in the noise-only category, deviated from the typical outcome data. This discrepancy arises from the testing metric's inability to convey meaningful information when accuracy scores reach 0%.

The percentage accuracy scores for samples 13 and 14 and the performance of the noise-only category were anomalies in the result data. This is because the testing metric does not account for a percentage accuracy of 0% meaningfully.

An example of an audio sample through all the stages is given to improve clarity. First, sample 14 is represented as a nearly silent audio sample, as indicated by the amplitude graph **(Figure 3-A)**. Subsequently, the amplitude graph of the added noise is represented **(Figure 3-B)**. The application of the stationary spectral subtraction algorithm for noise reduction is depicted also depicted along with the result obtained from the non-stationary spectral subtraction algorithm **(Figure 3-C, 3-D)**.

DISCUSSION

Our investigation aimed to address the question of how stationary and non-stationary spectral subtraction algorithms differ in their ability to reduce noise. By evaluating their performance across various noise categories, we gained insights into their respective strengths and limitations.

In the analysis of each noise category, we observed distinct patterns in the performance of the algorithms.

Audio samples with human speech and synthetic noise comprised the 'speech only' category. The non-stationary spectral subtraction algorithm performed better on average than the stationary spectral subtraction algorithm **(Figure 1)**. This result was as expected due to the dynamic nature of noise gates used in non-stationary spectral subtraction. The non-stationary algorithm takes into account consistencies and variations across the length of the sample to identify noise. The adaptability of the non-stationary algorithm allows it to better handle time-varying noise characteristics, which can be advantageous in dynamic noise conditions.

The 'speech with natural noise' category is the noise category most representative of real-life noise, and thus the performance of the algorithms in this category was of particular importance. As in the 'speech only' category, the non-stationary spectral subtraction algorithm performed better on average than the stationary spectral subtraction algorithm.

In the 'music with speech' sub-category, the non-stationary

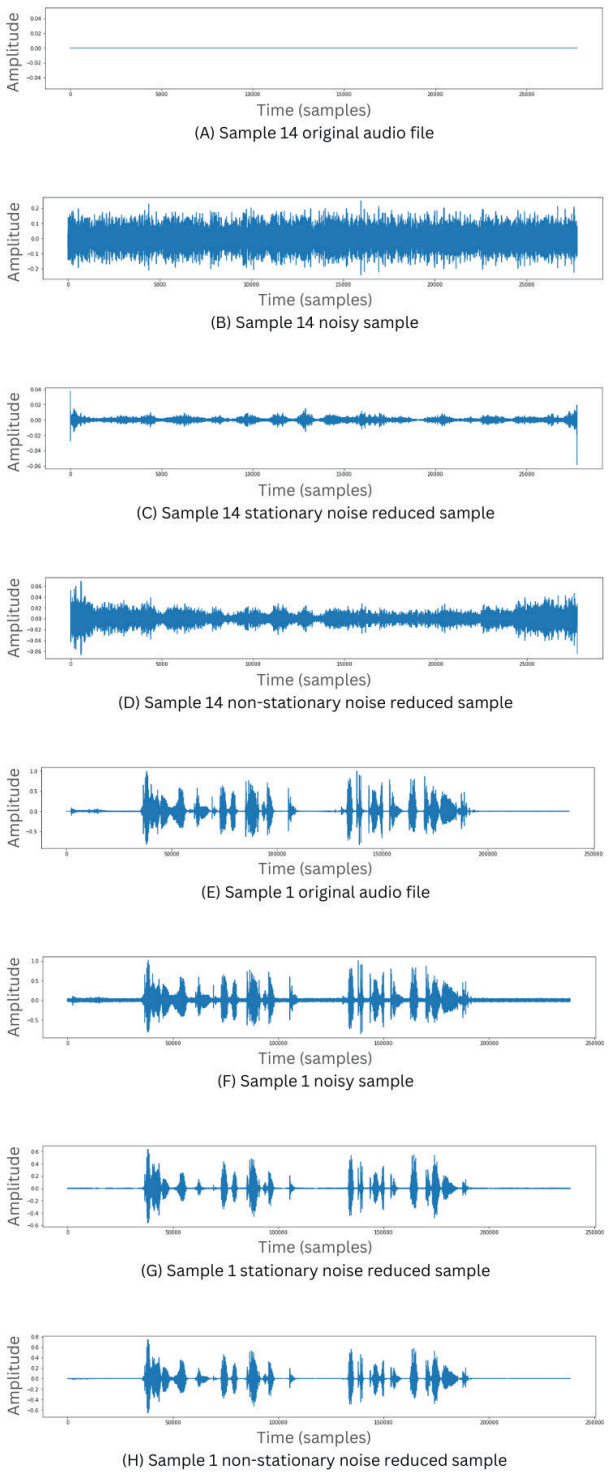

Figure 3: Sample 14 (A-D) and Sample 1 (E-H) through various stages. Sample 14 is initially a near-silent audio clip, evident from its amplitude graph (A). The graph for the added noise's amplitude is shown next (B). We present both the use of the stationary spectral subtraction algorithm and the outcome of the non-stationary spectral subtraction algorithm for noise reduction (C, D). In contrast, we offer a standard example to demonstrate noise reduction. The original sample 1 is displayed, followed by sample 1 after synthetic noise introduction (E, F). Results from applying the stationary spectral subtraction algorithm and the non-stationary spectral subtraction algorithm for noise reduction are also shown (G, H).

spectral subtraction algorithm performed significantly better than the stationary spectral subtraction algorithm. This suggests that in more complex sound mixtures, its ability to adjust the noise gates over time gives it greater advantage.

Stationary spectral subtraction performed slightly better than non-stationary spectral subtraction when denoising samples from the 'music without speech' sub-category. It can be inferred that the non-stationary noise reduction algorithm may have confused music with noise, since it assumes persisting audio patterns to be noise. The stationary noise reduction algorithm, on the other hand, gates the frequency bands indifferent to patterns in the sample, bypassing the possible confusion.

In the 'animal sounds' category, the stationary spectral subtraction algorithm performed significantly better than the non-stationary spectral subtraction algorithm. The anomaly in performance between the two algorithms in this category may have arisen because the algorithms were trained to primarily denoise human speech, the absence of which may have caused the drop in performance of the non-stationary spectral subtraction algorithm. Still, how the stationary noise reduction algorithm coped with the anomalous samples despite the same training remains inconclusive.

In analysing the results, we observed an anomaly in the noise-only category, specifically in samples 13 and 14, where one or more algorithms failed to produce a result. This anomaly can be exemplified by considering the graphs of sample 14 at different stages **(Figure 3**).

As outlined in the testing section earlier, the percentage accuracy score used is a normalized cross-correlation measure. As evident from the graphs, the algorithms were unable to achieve perfect noise reduction for sample 14. However, due to the comparison being made against a completely blank sample, even the slightest discrepancies significantly reduced the accuracy score, ultimately resulting in a score of zero percent.

In contrast to this anomalous case, we present a typical example showcasing the noise reduction process. The original sample 1 is illustrated, along with sample 1 after synthetic noise was added **(Figure 3-E, 3-F)**. The outcome of noise reduction using the stationary spectral subtraction algorithm and the result obtained from the non-stationary spectral subtraction algorithm is also depicted **(Figure 3-G, 3-H)**.

In conclusion, it was found that non-stationary spectral subtraction performed better in samples where human speech was the target: speech only, speech with natural noise and music with human speech, while stationary spectral subtraction performed marginally better when denoising music only, and significantly better when denoising animal sounds.

Despite the anomaly in objective results, tests like the manual comparison of graphs and listening to the sample established that stationary noise reduction performed better with the last category: noise only.

The quantitative test used in the procedure was based on a comparison to the 'ideal' noise-reduced sample. This was assumed to be the original sample S_o since denoising a synthetically noised sample was expected to return the original input. The synthetic noise generated may not perfectly represent real-world noise. Still, the noise reduction performance measure should hold value.

However, in some categories, the original sample may

have contained noise prior to the addition of synthetic noise. This may have affected the reliability of the similarity test with the speech with natural noise, music without speech and noise-only categories in particular.

The noise, when synthetically added, did not vary with time. This is a limitation to the representativeness of the sample space, since in the real world, there may be scenarios where noise varies significantly with time. In future studies, the noise can be generated more organically, rather than using white noise, to better test the robustness of these techniques.

In this study, we use two types of noise reduction techniques. Comparison to other common methods of noise reduction, *e.g.,* the Butterworth and Wiener Filtering, would prove useful.

A profound application of spectral subtraction noise reduction could be in its use in hearing aids. By distinguishing speech from noise, the speech-to-noise ratio SNR could be enhanced, aiding the perception and comprehension of speech in the presence of background noise (8).

MATERIALS AND METHODS

An overview of the process if illustrated **(Figure 4)**. The first step in the process is the decomposition of the sound sample into frequency bands for the second step: the noise threshold calculation for each frequency band (6).

Sound is processed by microphones, and consequently computers, in terms of an intensity-time graph. Sound is a mostly periodic function of pressure differences in a medium (the intensity factor) with respect to time. Consider the simple periodic signal:

$$f(t) = \cos(\pi t)$$

This signal represents a sound consisting of a pure cosine wave with a frequency of 0.5 (time period 2).

Sound in the real world, however, consists of a multitude of such pure signals. It is difficult to, therefore, apply frequency-related operations to it. To decompose a complex sound to its constituent frequencies, Fourier transformation is used (9, 10). Consider the following compound signal:

$$g(t) = \cos(\pi t) + \cos(4\pi t)$$

This compound signal is a combination of two primary signals with their own respective frequencies of 0.5 and 2.

$$g^*(f) = \int_{t_1}^{t_2} g(t) e^{-2\pi i f t} dt$$

The Fourier transform g^* is the amplitude-frequency representation of the original intensity-time signal g. The domain of the signal is converted from time (t) to frequency (f).

The value of the Fourier transforms g^* oscillates around 0 for most of the frequencies but shows a spike at 0.5 and 2, thus decomposing the signalg into its constituent frequencies (10).

In this manner, the complex audio signal is decomposed into its various frequencies and categorized into frequency bands, the noise gates for which are then calculated.

The application of the gate is essentially the removal of the estimated noise signals (noise gates) from the compound signal. The Fourier Transforms of the noise and the noisy signal are 'subtracted' from each other. The resulting Fourier Transform is then processed through the inverse Fourier transform, converting a signal in the frequency domain (f) back to its corollary in the time domain (g) (6).

Sample Testing on Spectral Subtraction Algorithms

The procedure for testing a sample on either of the algorithms includes three primary steps. An original audio file (S_O) in WAV format was uploaded to GitHub. The algorithm accessed this file as GitHub user content.

Synthetic noise was added to each S_O to obtain the noisy sample (S_N). The synthetic noise was a blend of sounds of various frequencies and the configuration of this noise was dependent on 3 arguments: minimum frequency, maximum frequency, and intensity factor. The noise varied from sample to sample to ensure a more representative sample space. The addition of noise was a plain overlapping of both waveforms. S_N was then passed through stationary and/or non-stationary noise reduction to obtain the noise-reduced audio S_R. Both algorithms were constructed and run in a Python environment.

Testing

If S_O is a noiseless audio file, the ideal outcome of noise reduction should be the same as S_O. This means that whatever synthetic noise was added to S_O was completely removed, keeping the desired sound intact.

Based on the ideal case, the two algorithms were tested on a percentage similarity figure of S_O and S_R. Percentage accuracy is the normalised cross-correlation between the

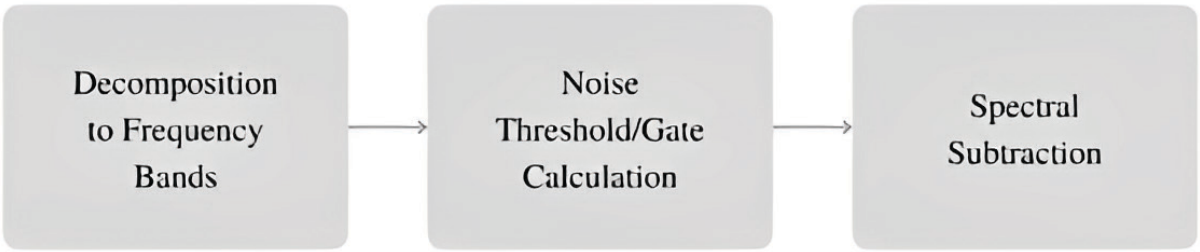

Figure 4: Overview of steps undertaken in the spectral subtraction algorithm. The first step in the process is the decomposition of the sound sample into frequency. Using Fourier transforms, the complex audio signal is decomposed into its various frequencies and categorized into frequency bands. Then noise gates are calculated for the various frequency bands. Finally, the Fourier transforms of the noise and the noisy signal are 'subtracted' from each other, in a process called spectral subtraction.

Noise Category	Sample size	Original audio (S_O)	Noised audio (S_N)	Description
Speech only	4	Noise-free human speech.	Synthetic noise was added.	This category involves noise reduction algorithm testing using human speech without significant background noise, serving as a basic test scenario.
Speech with natural noise	3	Noisy human speech.	Synthetic noise was added to selected samples.	This category simulates real-world applications by incorporating human speech along with ambient background noise during recording.
Music	4	Samples including any form of music.	Synthetic noise was added to selected samples.	This category encompasses samples containing a combination of music and speech, or solely music, to evaluate the algorithms' ability to differentiate between music and noise. It also tests their capability to prioritize the audio subject.
Animal sounds	2	Samples including any form of animal sounds.	Synthetic noise was added.	This category focuses on samples featuring animal sounds, as separating animal sounds from noise is crucial when the animal sound is the subject. Similar to the previous category, it assesses the algorithms' subject prioritization skills.
Noise only	2	Either natural noise, or completely silent audio.	Synthetic noise was added to the silent samples only.	The last category involves more experimental samples consisting of pure noise, contributing to a more exploratory aspect of the study.

Table 2: Description of the composition of the different noise categories.

noise-reduced audio S_R and the original recording S_O. Effectively, it is a score of how similar the noise-reduced audio is to the ideal noise-less audio.

Apart from percentage accuracy, the graphs and audios of the S_O, S_N, and S_R were also recorded and studied (11).

Statistical tests were conducted using the online statistical tool DATAtab. The analysis involved the addition of values and the application of the Mann-Whitney U test, which is a non-parametric test suitable for comparing two independent groups. In this study, a significance level of 0.05 was chosen as the threshold to determine statistical significance.

Data Collection for the Algorithm

The 15 data samples collected for the purpose of testing the algorithms were divided into 5 broad categories. The primary 3 categories are relatively noise-free common sounds with added noise.

Note that in some categories, synthetic noise was added selectively to some samples only, since many already had natural noise. This also ensures that a more diverse and representative range of noise is included.

ACKNOWLEDGMENTS

We would like to acknowledge Ms. Kah Ying for help with writing, and Ashwin Malik for constructive feedback on this work.

Received: October 24, 2022
Accepted: April 18, 2023
Published: September 20, 2023

REFERENCES

1. The Chicago School of Media Theory. "Noise." Lucian.uchicago.edu.
2. “Noise Pollution: A Modern Plague.” Noise Pollution: A Modern Plague, docs.wind-watch.org/goineshagler-noisepollution.html.
3. Selesnick, I. W., and Burrus, C. S. "Generalized digital Butterworth filter design." IEEE Transactions on Signal Processing, vol. 46, no. 6, June 1998, https://doi:10.1109/78.678493.
4. Guduguntla, Praneeth. "Background Noise Removal: Traditional vs AI Algorithms." Medium.
5. Park, Se Rim, and Lee, Jin Won. "A Fully Convolutional Neural Network For Speech Enhancement." Arxiv.org.
6. Valin, Jean-Marc. "RNNoise: Learning Noise Suppression." Jmvalin.ca.
7. Sainburg, T. "timsainb/noisereduce: v1.0 [Software]." Zenodo, 2019, doi: https://doi.org/10.5281/zenodo.3243139.
8. National Guideline Centre (UK). "Hearing aid microphones and noise reduction algorithms." Ncbi.nlm.nih.gov.
9. Paialunga, Piero. "Noise cancellation with Python and Fourier Transform." Medium.
10. MathWorks. "Fourier Transforms." Mathworks.com.
11. Silva, Thalles S. "Practical Deep Learning Audio Denoising." Sthalles.github.io.