Efficiency of Spectral Subtraction Algorithms for an Urban Audio Acquisition System Using IoT Devices

Evan Fallis, Petros Spachos and Stefano Gregori
School of Engineering, University of Guelph, Guelph, Ontario, Canada

***Abstract*—Ambient audio acquisition systems gather information about the noise levels in a specific area. Such systems chart the environment based on averaged sound parameters as they change over time. This can improve human life by detecting nuisance noise as well as sound levels hazardous to human health and by allowing citizens, business owners and urban planners to visualize the noise pollution in their city. An efficient way to monitor the noise is through cost-effective, small-size Internet of Thing (IoT) devices that monitor, record and report the noise level. However, the prevalence of wind and electrical interference and artefacts in urban environments affects the fidelity of audio acquisitions. Filtering out these unwanted contributions allows the analysis of relevant sounds. In this paper, three spectral subtraction algorithms are compared and applied to urban sound clips to reduce unwanted noise from audio recordings. The effectiveness of the Boll, Berouti, and Kamath algorithms for reducing noise in various sounds found in urban environments was assessed. Then, an IoT device was used to collect real data from an outdoor environment, using a spectral subtraction algorithm. According to experimental results, the Kamath algorithm increases the signal-to-noise ratio by 20 dB and improves the resemblance to the true audio signal.**

***Index Terms*—Noise pollution, acoustic noise, spectral subtraction.**

I. Introduction

Smart cities target the needs of the public and focus on systems that are beneficial for sustained health and well-being of the citizens. The collection of environmental audio data allows the public to visualize their city based on noise pollution levels. It can also help mitigate noise to reduce the amount of physical and mental damage from long exposures to loud noises [1]. The overall cost, the ease of deployment and the accuracy of such systems are challenges that need to be properly addressed to make them viable [2]. Internet of Thing (IoT) devices can alleviate these problems. The intention is to let the public visualize the noise they experience in different areas and at different times. However, removing noise not perceived by humans and stray artefacts is a challenge when considering ambient audio collection systems. Filtering out the noise that is not relevant, such as wind, mechanical vibrations, or electrical interference, is important to produce a clean signal which represents how a human would perceive the environment. Typical speech enhancement algorithms may filter out noises that are intended to be kept such as a dog bark or a vehicle engine.

In terms of quantifying results, there are multiple ways to show the quality of a signal numerically. Popular methods include the Signal-to-Noise Ratio (SNR) and Perceptual Evaluation of Speech Quality (PESQ) [3], [4]. Unfortunately, these methods are not always enough to measure the effectiveness of an algorithm since they do not reference the ideal output, hence, an unwanted output signal could still be considered high quality. Finding the difference between the ideal output and the actual one will allow for an objective way to assess the accuracy resulting from each algorithm. This will also show any improvements when comparing the outputs to the original signal before filtering.

In this work, three spectral subtraction techniques are analyzed and used to reduce unwanted noise in multiple urban city sound clips. This includes the Boll, Berouti, and Kamath algorithms. A comparison of the three algorithms was conducted through experiments, to find the most efficient in terms of signal quality for urban sounds. A collection of urban noises which has four main urban sound categories was used for the comparison. The results are analyzed based on the SNR, magnitude-squared coherence, and time overhead of each algorithm. Then, an IoT device is used to collect real data from an outdoor environment and the results are discussed.

The rest of the paper is organized as follows: Section II includes the related work followed by Section III with a brief description of each of the algorithms used. Section IV presents the system framework, and Section V discusses the results. The conclusion is in Section VI.

II. Related Work

Spectral subtraction to remove unwanted noise from voice recordings has been studied in the literature [5]–[8]. In [6], an experiment targeting the patterns made with vowels in the frequency domain was conducted. This limits the applicability of the method since vowels are specific to human voice, not necessarily applicable to all urban sounds. In [7], a noisy signal of human voice was improved by 1.72 dB after using the multi-band spectral subtraction method. In [8], an approach to noise removal using the concept of short term energy was discussed. It adapts the signal based on the changing levels of noise in the environment rather than assuming all noise is static. In [9], a modified spectral subtraction algorithm designed to clean a speech signal was developed. The objective parameter being compared was the SNR of the input and output signal. An additive white Gaussian noise channel was used to simulate noise in the speech signals. In [10], a spectral subtraction algorithm for detecting speech emotions was proposed. The technique was shown to significantly reduce unwanted noise when attempting emotion recognition.

Several works compare the effectiveness of spectral subtraction algorithms [11], [12]. In [13], four popular spectral subtraction algorithms were compared. The findings show that there is no clear winner when considering all conditions. In [14], different forms of spectral subtraction were analyzed, using SNR and PESQ as objective measurements. Although SNR and PESQ both measure the quality of a signal, they do not show how similar the output is when compared to the target signal. In [15], methods of spectral subtraction by recording the SNR are considered. Objectively quantifying audio quality is important when analyzing the effectiveness of signal processing algorithms. In [16], an objective measurement for ambisonic spatial audio is developed, which is capable of predicting the listening quality and localization accuracy of audio by using ambisonics. In [17], an objective measurement of audio quality based on advanced audio coding compressed domain is studied using a score validated by comparing it to a subjective rating, showing a clear trend of the objective measurement coinciding with the subjective appraisal.

In comparison to the related work, this paper addresses the concept of using spectral subtraction for ambient audio collected in urban environments. The challenge is that urban sounds are not exclusively human speech. Certain existing algorithms would filter out useful sounds in urban environments. Observing and analyzing which algorithms work best for sounds found in urban environments bring smart cities closer to reality. In contrast to other works, this paper validates the use of spectral subtraction when applied to urban sounds.

III. Spectral Subtraction Algorithms

Spectral subtraction refers to sampling noise in an audio clip and subtracting the result from the entire recording in the frequency domain [18]. In this work, we examine the performance of three spectral subtraction algorithms, Boll [18], Berouti [19], and Kamath algorithm [20]. The algorithms are selected due to their simplicity and low complexity to be implemented in an IoT device with limited energy requirements. They are compared in terms of their efficiency for an IoT urban noise collection system.

A. Boll Algorithm

The Boll algorithm was one of the original works referencing the technique of spectral subtraction. The technique is described as:

$$S(k) = [|X(k)| - B(k)]\, e^{jX(k)}, \tag{1}$$

where $S(k)$ is the resulting signal, $X(k)$ the input signal, $B(k)$ the average noise spectral magnitude, and k the frequency.

The intention is to target speech activity and remove any unrelated sound. This is done by targeting a segment of a clip with no active speech and converting that segment into the frequency domain. The entire signal is then converted to the frequency domain and the spectral energy of the segment is removed from the signal. Boll algorithm was selected due to its simplicity and easy implementation in a cost-effective IoT device.

B. Berouti Algorithm

The Berouti algorithm focuses on removing noise in speech. The algorithm is illustrated as:

$$D(w) = P_s(w) - \alpha P_n(w) \tag{2}$$

where P_s is the spectral signal power, and w the frequency. This is defined as:

$$P_s' = \begin{cases} D(w) & D(w) > \beta P_n(w) \\ \beta P_n(w) & \text{otherwise} \end{cases} \tag{3}$$

$$\text{assuming } \alpha \geq 1 \text{ and } 0 < \beta \ll 1$$

where P_n is the spectral power of the noise, α is the subtraction factor, and β represents the spectral floor.

This algorithm was intended to extend the Boll theory by removing anomalies created in the original algorithm. This was done by altering the method in two ways. First, the noise power is overestimated. Second, the spectral power of any segment cannot be reduced below a certain threshold, known as the spectral floor. These anomalies are typically observed through listening to the filtered audio. Since the target of the proposed system is to improve audio quality from a technical point of view rather than that of human perception, the extension proposed by Berouti does not necessarily help in terms of objective analysis.

C. Kamath Algorithm

The Kamath algorithm is an adaptation of the Berouti algorithm. Its purpose is to reduce the amount of distortion caused by a traditional method while maintaining its effectiveness. This algorithm is defined as:

$$|\hat{S}_i(k)|^2 = |Y_i(k)|^2 - \alpha_i \sigma_i |\hat{D}_i(k)|^2 \quad b_i \leq k \leq e_i \tag{4}$$

where Y_i is the original signal, D_i a sample of the noise, α_i the over-subtraction factor, σ_i the tweaking factor, b_i the beginning of the frequency bin, and e_i the ending of the frequency bin, all with respect to the i-th frequency band.

This algorithm revolves around the idea that real-world noise is typically not white. Whereas white noise has a constant spectral power regardless of frequency, colored noise has a non-uniform spectral distribution. Hence, this algorithm uses a multi-band technique that takes into account the variations of real-world colored noise. Similar to Berouti, the main performance metric was human perception.

IV. System Framework

A system was designed for experimentation to examine the efficiency of each spectral substitution algorithm, when IoT devices are used for audio collection and processing. A system overview is shown in Fig. 1. Selected sound clips from a labeled dataset are played from a speaker and recorded by the microphone of an IoT device. The distance between the speaker and the microphone was within 5 m. The data from the microphone are forwarded to an edge device for processing and classification. The edge device communicates also with the cloud for data classification and storage.

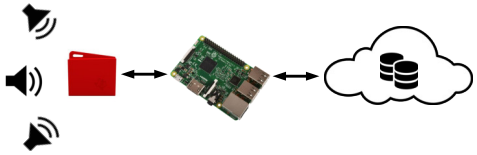

Fig. 1: Overview of the system framework.

The experiment took place in an outdoor environment at a university campus with people and vehicles moving around. Each spectral subtraction algorithm was applied to the collected signal to improve its quality. The SNR was calculated as well as a coherence comparison with the original signal to check the similarity between the two signals. The delay contributed to the script based on the chosen spectral subtraction algorithm was also measured.

A. Audio Collection and Processing Devices

To collect audio data, the Texas Instruments SensorTag was used, as an IoT device with a microphone. SensorTag contains a variety of sensors that are controlled via the microprocessor. It uses Pulse Density Modulation (PDM) to transfer data from the microphone to the microprocessor. Code Composer Studio (CCS) was used to program the SensorTag in order to enable the microphone and disable all other services that were not related to the operation of collecting and transferring audio data. This is necessary to save energy and extend the lifetime of each device.

SensorTag is powered with a coin cell battery. The typical capacity of the battery is 235 mAh at 2 V, which corresponds to approximately 470 mWh. The capacity can be divided by the average power to find the estimated lifetime, as seen in the general formula:

$$\text{Battery Lifetime} = \frac{\text{Total Capacity}}{\text{Average Power}} \tag{5}$$

The power usage of the SensorTag when streaming audio data continuously at the maximum rate is shown in Fig. 2. The spikes in power are from both the sampling of audio and the transmission of each frame. Following the energy requirements, Table I shows the estimations of battery life. Increasing the delay between samples reduces the average power consumption to less than the minimum of the default, which can be seen in Fig. 2. This shows the effectiveness of decreasing the sampling rate in terms of energy consumption.

The audio data from the SensorTag are forwarded over Bluetooth to an edge device for further processing. A Raspberry Pi 3 is used as an edge device to collect the audio data fed by the microphone.

B. Sound Acquisition

To simulate the noise in a realistic environment, the audio clips were played using high-quality loudspeakers. The ST MP34DT05-A MEMS microphone was used to collect all audio clips. The microphone samples at a rate of 44.1 kHz and has an SNR of 64 dB(A). It was connected to the ST X-NUCLEO-CCA02M1 microphone shield, an interface

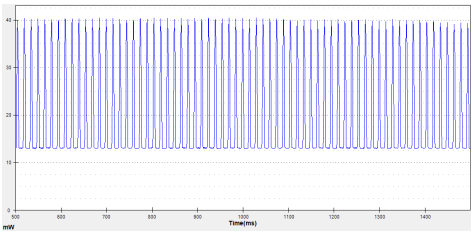

Fig. 2: Power consumption of SensorTag.

TABLE I: Estimation of Power Consumption

Delay (ms)	Power (mW)	Life (days)
0	19.77	0.99
50	9.14	2.14
200	7.66	2.56

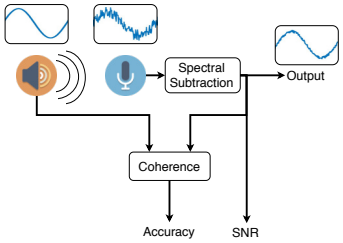

Fig. 3: Overview of the signal processing elements.

designed to interpret PDM signals coming from the digital microphone. The shield is capable of interfacing through USB with a personal computer. Finally, it was connected to an ST NUCLEO-F401RE microcontroller, which provided the appropriate processing. Audacity, free software designed to have many functions relating to audio, was used to record by using the aforementioned components.

C. Audio Clip Selection

Audio clips were selected from the Urban Sounds dataset [21]. This dataset contains four categories that are meant to represent the variety of sounds found in urban cities. The categories include human, music, mechanical, and nature. All sounds chosen were 4 s long, the standard slice length for the dataset. This is a sufficient length when classifying clips in the original urban sound work [21]. Ten clips from each category were used to provide a variety in terms of audio.

D. Evaluation Metrics for the Spectral Subtraction Algorithms

The overview of the signal processing components during experimentation is shown in Fig. 3. Three metrics are used: SNR, magnitude-squared coherence, and clip duration.

1) SNR: The SNR measures how noisy the output signal is. It shows the signal power when compared to the noise, hence the higher the value the better [3]. The purpose is to measure the quality of the signal rather than comparing it to any other signal. The general equation for SNR in dB can be seen below:

$$\text{SNR}_{\text{dB}} = 10 \log_{10} \frac{P_{signal}}{P_{noise}} \tag{6}$$

where P_{signal} is the spectral power that composes the signal and P_{noise} is the spectral power of the noise. Since the power of the signal component after each algorithm is unknown, it can be derived using the average noise at a point which only contains noise. This can be expressed as:

$$P_{signal} = P_{total} - P_{noise} \tag{7}$$

with P_{total} being the spectral power of the entire signal with no alterations. This allows a substitution into (6) which results in the following:

$$\mathrm{SNR_{dB}} = 10 \log_{10} \frac{P_{total} - P_{noise}}{P_{noise}} \tag{8}$$

meaning that the SNR can be calculated as long as a point with only noise is identifiable in the clip.

The estimated noise after each algorithm was estimated and kept consistent for each clip. However, this alone does not show how effective an algorithm is, because it is not compared to the ideal signal, the one being played from the loudspeakers in our case. It does not necessarily represent accuracy since the ratio only considers the spectral powers rather than how closely the signal represents the original. This can be done using the magnitude-squared coherence, a method of analysis to measure the overlap of frequencies present.

2) Magnitude-Squared Coherence: The filtered audio clips need to be compared to the source. This was done via coherence, which compares the magnitude of frequencies in the signal, and is defined as:

$$C_{xy}(f) = \frac{|P_{xy}(f)|^2}{P_{xx}(f)P_{yy}(f)}, \tag{9}$$

where P_{xy} represents the cross-spectral density given by:

$$P_{xy}(f) = \int_{-\infty}^{\infty} R_{xy}(t) e^{-j\omega t} dt, \tag{10}$$

where R_{xy} is the cross-correlation of the two original signals, $x(t)$ and $y(t)$. The cross-correlation of both signals can be seen below:

$$R_{xy}(t) = \overline{x(-t)} * y(t) \tag{11}$$

where the two signals are convoluted to measure the overlap. These equations demonstrate how an overlap in the frequency domain can be accurately measured using the standard coherence procedure.

3) Delay: Another useful metric is knowing the delay or duration of time needed to perform each algorithm, especially when IoT devices with limited energy resources are used. This also helps estimate the delay overhead when implementing a spectral subtraction algorithm. Although the impact may be small if considering the delay of a single use of a spectral subtraction algorithm, repeated uses on multiple audio clips can stack up and create a significant delay.

The delay was calculated by measuring a Matlab implementation of each algorithm. The simulation was run using a 3.6 GHz Intel Xeon processor, 16 GB of RAM, on 64 bit Windows 10 Pro. The data was collected by averaging 10 runs of each algorithm for every category. Each use of the algorithm was conducted on a 4 s clip, at a sampling rate of 96 kHz, totaling to an array size of 384,000 data points.

V. Experimental Results and Discussion

The dataset contained four categories and each of the four categories was used with all three algorithms tested. The results are from the twelve experiments performed. The analysis was performed in Matlab after collecting the data. The scripts were uploaded for public use at [22].

A. SNR Analysis

A summary of the results relating to the SNR can be seen in Fig. 4. Any modification being done to the signal, starting with no alteration and followed by the three algorithms being applied, is shown along with the estimated SNR value. It is clear that the SNR is improved significantly in every case when an algorithm is applied. This is because the frequencies related to noise have been subtracted from the clip, with slight variations for each algorithm. Kamath and Boll have similar performance, while all the three approaches have their highest performance for the music category.

The most noticeable difference between the applied algorithms is that Berouti is slightly worse when it comes to increasing the SNR of a signal. However, a signal can be accurate even if the SNR is not high, but it is higher than a threshold. As it can be inferred, one prerequisite to having a good signal is having a sufficient SNR. The applied algorithms show that they have this prerequisite based on the SNR levels.

B. Frequency Analysis

A simple correlation between the two signals will not always show the wanted result because frequencies are more important than amplitudes when dealing with audio signals. Analyzing how closely the frequency domain of the two signals match will indicate how closely they are related.

The coherence of the two signals will show the overlap of frequencies common between both signals. The accuracy found for the coherence is shown in Fig. 5. For every category, each algorithm increases the accuracy by a significant amount. Since the original target for spectral subtraction algorithms was human voice, it was expected that each algorithm would consistently increase the overall accuracy. For mechanical noise, the accuracy after each algorithm increased the least compared to other categories. This is most likely because mechanical sounds are more periodic than others, a feature that spectral subtraction specifically tries to remove.

Music had the most positive impact, because typically has more periodic frequencies due to its deliberate nature. Any noise in the clip can easily be distinguished from the signal. This explains why the music category would be positively impacted. Finally, nature was positively impacted when using each algorithm. Nature sounds often have less frequent activity, such as crickets or a bird chirp. This leads to much more noise being present than signal, meaning that spectral subtraction should be very effective. In terms of effectiveness, all algorithms significantly increased the coherence accuracy.

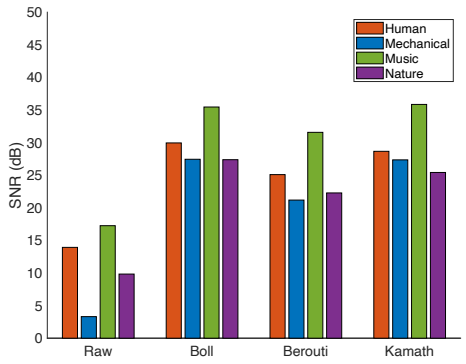

Fig. 4: SNR results.

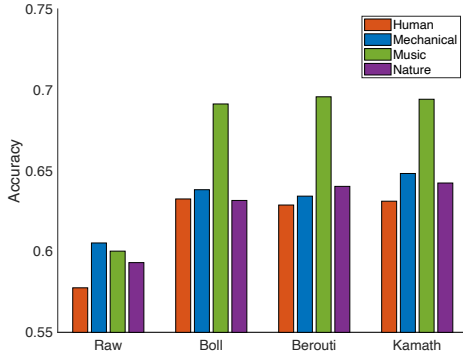

Fig. 5: Coherence results.

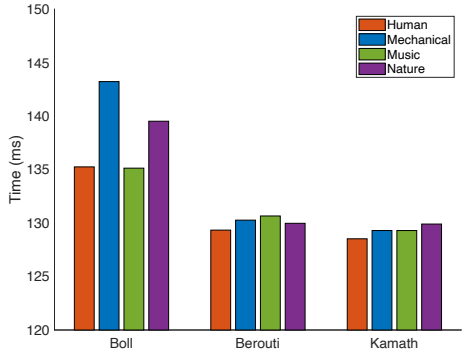

Fig. 6: Algorithm delay results.

There is a small difference in each algorithm, the main difference being the slightly higher accuracy of Kamath in every category.

C. Algorithm Delay Analysis

The delay of each algorithm helps indicates how well it can be implemented into a real-time system. It also determines how fast an analysis of multiple audio clips will be. The results of this experiment can be seen in Fig. 6. Both Berouti and Kamath took approximately 130 ms on average. Although this is very low, this is important for a low-power system.

Boll algorithm took from 135 to 145 ms depending on the type of sound. Interestingly, the mechanical and nature categories took slightly longer than others when using Boll. This may be explained by the frequencies present, both mechanical and nature have high frequencies which may add a small overhead specifically with Boll. Thus, Boll has a slight disadvantage when compared to others if the timing is a priority. This does not necessarily apply to all systems since an individual implementation may be able to improve the time complexity.

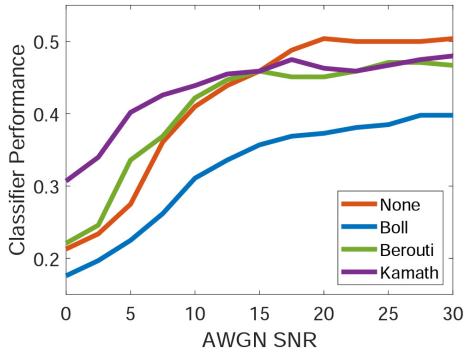

Fig. 7: Classifier performance.

D. Classifier Performance Comparison

The classifier performance based on filtering is shown in Fig. 7. A range of tests run with varied SNRs in the AWGN. It can be seen that both Berouti and Kamath improve the performance of the classifier when there is much more noise, Kamath being more significant than Berouti. This improves performance by 10% when considering a noisy signal.

The inconsistency based on SNR makes sense as the effectiveness of spectral subtraction exhibits diminishing returns when the signal already has a low level of noise. It would be useful to arbitrate based on the SNR of the data collected to determine whether a spectral subtraction algorithm should be applied. Spectral subtraction could also be used when the initial classifier confidence is low, filtering and running the classifier again could improve the accuracy.

E. Waveform Processing

A comparison of the waveform after using simple filters and processing is shown in Fig. 8. It is important to note that the waveform being used is sampled at full speed, meaning that the power savings technically do not apply to this scenario. The processing of the signal would be the same if the artificial delay between samples was present. The raw waveform shows the basic values read from the .wav file. The average is close to zero as it considers both the positive and negative magnitude. The absolute waveform simply takes the raw data and converts it to positive numbers, then calculates the average of 0.0016. The envelope filter attempts to track the magnitude of the waveform, this instance being slow. The average was 0.0019, slightly higher than the previous.

The purpose of this experiment is not to calculate the actual decibels measured, but rather to show how this technique can be used to find the relative noise level and then normalize the magnitude in later experiments. Storing the raw data on a server is typically a good technique since it can be hard to obtain the original after signal processing has taken place.

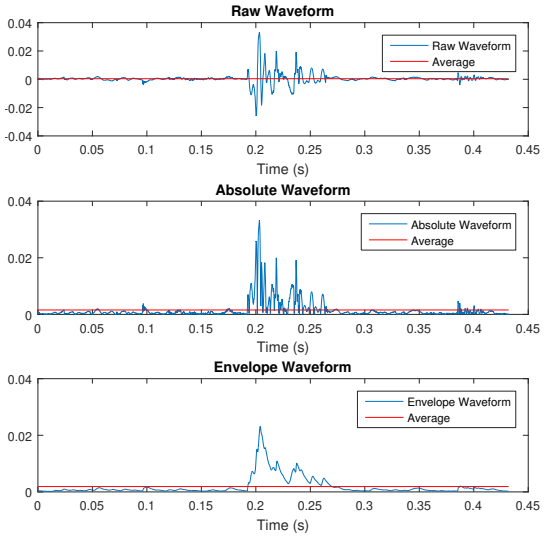

Fig. 8: Envelope analysis.

F. Discussion

It is important to understand the impact of the results and what this might imply. As discussed before, the SNR does not show much aside from the quality of the signal. Having a sufficient value is required for the signals to match. As shown by the results, Kamath achieved the highest accuracy while having a sufficiently high SNR. To reiterate, Kamath considers colored noise rather than white. This makes it better since the real world does not equally weight frequencies. Although the original performance metric of the Kamath approach was a subjective test, the accuracy of the output signal was seen to be better due to the multi-band approach. The discussed factors explain why Kamath was seen to be better overall. In terms of implementing each algorithm into a system, Kamath shows the best results, showing the maximum accuracy achievable of the tested algorithms and low energy requirements. Berouti is a good choice even when compared to Kamath. The results are close to Kamath despite being a slightly different approach. Boll algorithm shows the least benefits, the only difference being the simplistic nature compared to the others. Despite this, it has the highest delay.

VI. Conclusion

In this paper, three spectral subtraction algorithms were tested with various types of urban sounds. The objective measurements included SNR, coherence accuracy, and delay. The Kamath algorithm was found to have the most positive impact, increasing the SNR by up to 20 dB while also increasing the accuracy significantly. The music category of the Urban Sounds dataset was seen to have the biggest positive overall impact. The Kamath algorithm had the least delay during the spectral subtraction process it was finally selected and implemented in an IoT device. During real experimentation, the energy requirements of the algorithm when implemented in the IoT device are promising.

References

[1] D. Shepherd, K. Dirks, D. Welch, D. McBride, and J. Landon, "The covariance between air pollution annoyance and noise annoyance, and its relationship with health-related quality of life," *International Journal of Environmental Research and Public Health*, vol. 13, no. 8, 2016.

[2] J. Salamon and J. P. Bello, "Deep convolutional neural networks and data augmentation for environmental sound classification," *CoRR*, vol. abs/1608.04363, 2016.

[3] D. R. Pauluzzi and N. C. Beaulieu, "A comparison of snr estimation techniques for the awgn channel," *IEEE Transactions on Communications*, vol. 48, no. 10, pp. 1681–1691, Oct 2000.

[4] A. W. Rix, "Comparison between subjective listening quality and p. 862 pesq score," *Proc. Measurement of Speech and Audio Quality in Networks (MESAQIN'03), Prague, Czech Republic*, 2003.

[5] E. Fallis, P. Spachos, and S. Gregori, "A power-efficient audio acquisition system for smart city applications," *Internet of Things*, vol. 9, p. 100155, 2020.

[6] C. Lu, K. Tseng, Y. Chen, L. Wang, and C. Lei, "Speech enhancement using spectral subtraction algorithm with over-subtraction and reservation factors adapted by harmonic properties," in *2016 International Conference on Applied System Innovation (ICASI)*, May 2016, pp. 1–5.

[7] J. C. Saldanha and Shruthi O R, "Reduction of noise for speech signal enhancement using spectral subtraction method," in *2016 International Conference on Information Science (ICIS)*, Aug. 2016, pp. 44–47.

[8] S. S. Bharti, M. Gupta, and S. Agarwal, "A new spectral subtraction method for speech enhancement using adaptive noise estimation," in *2016 3rd International Conference on Recent Advances in Information Technology (RAIT)*, March 2016, pp. 128–132.

[9] S. S. Meher and T. Ananthakrishna, "Dynamic spectral subtraction on awgn speech," in *2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN)*, Feb 2015, pp. 92–97.

[10] Z. Wanli, L. Guoxin, and W. Lirong, "Application of improved spectral subtraction algorithm for speech emotion recognition," in *2015 IEEE Fifth International Conference on Big Data and Cloud Computing*, Aug. 2015, pp. 213–216.

[11] Y. Hu and P. C. Loizou, "Subjective comparison of speech enhancement algorithms," in *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, vol. 1. IEEE, 2006, pp. I–I.

[12] R. Martin, "Spectral subtraction based on minimum statistics," *power*, vol. 6, p. 8, 1994.

[13] T. K. Dash, P. Rout, P. R. Dash, S. Sahoo, and A. Alkama, "Objective quality measures evaluation of the different spectral subtraction algorithms in various noise conditions," in *2018 3rd International Conference on Communication and Electronics Systems (ICCES)*, Oct 2018, pp. 143–146.

[14] N. Upadhyay and A. Karmakar, "Speech enhancement using spectral subtraction-type algorithms: A comparison and simulation study," *Procedia Computer Science*, vol. 54, pp. 574–584, 2015.

[15] M. Yektaeian and R. Amirfattahi, "Comparison of spectral subtraction methods used in noise suppression algorithms," in *2007 6th International Conference on Information, Communications Signal Processing*, Dec 2007, pp. 1–4.

[16] M. Narbutt, A. Allen, J. Skoglund, M. Chinen, and A. Hines, "Ambiqual - a full reference objective quality metric for ambisonic spatial audio," in *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, May 2018, pp. 1–6.

[17] W. Qian, Y. Zhao, and Z. W. Tu, "Objective audio quality evaluation methon based on compressed domain," in *2018 IEEE International Conference on Computer and Communication Engineering Technology (CCET)*, Aug. 2018, pp. 172–175.

[18] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, no. 2, pp. 113–120, 1979.

[19] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *ICASSP '79, IEEE*, vol. 4, April 1979, pp. 208–211.

[20] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise." in *ICASSP*, vol. 4. Citeseer, 2002, pp. 44 164–44 164.

[21] Urban Sound Data Sets. (retrieved: 2022-03-07). [Online]. Available: https://urbansounddataset.weebly.com/

[22] Matlab Scripts Repository. (retrieved: 2020-11-01). [Online]. Available: https://github.com/efallis/urbanSoundsSpectralSubtraction