COVERSHEET

THE UNIVERSITY OF WESTERN AUSTRALIA
Achieve International Excellence

Faculty of Engineering, Computing and Mathematics
Assignment, Report & Laboratory Coversheet for Individual & Group Assignment

SUBMITTING STUDENT		
SURNAME	**GIVEN NAMES**	**STUDENT NUMBER**
UNIT NAME SIGNAL PROCESSING		**UNIT CODE** ELEC4404
TITLE/TOPIC OF ASSIGNMENT SIGNAL PROCESSING GROUP PROJECT.		**NAME OF LECTURER/TUTOR** LU JIN
DATE/TIME DUE 1st June 11 am		**DATE/TIME SUBMITTED**

HONOURS STUDENTS ONLY	OFFICE USE ONLY
By signing this document, I further assert that the length (word count) of my dissertation is within the maximum allowed length governed by the project unit I am enrolled in. Penalties, as outlined on this website, will be applied for over length dissertations.	

FOR GROUP ASSIGNMENTS ONLY NAME	STUDENT NUMBER
1. Leon Kingsley	21305101
2. Euclid Kosasih	21315919
3. Khanh-Tan Phan	21326604
4. Luke Warner	20930781
5.	
6.	
7.	
8.	

Unless other arrangements have been made it will be assumed that all group members have contributed equally to group assignments/laboratory reports

DECLARATION

I/We are aware of the University's policy on academic conduct (see over) and I/We declare that this assignment/project is my own/my group's work entirely and that suitable acknowledgement has been made for any sources of information used in preparing it. I/We have retained a hard copy for my/our own records.

SIGN: Leon Kingsley	**SIGN:** Euclid K
SIGN: KP	**SIGN:**
SIGN:	**SIGN:**
SIGN: Warner	**SIGN:**

NOTE: No assignment will be accepted without the declaration above being signed and dated
SEE OVER FOR INFORMATION ON REFERENCING & PLAGIARISM

THE UNIVERSITY OF WESTERN AUSTRALIA

Achieve International Excellence

Signal Processing

ELEC4404

Forensic Analysis Project

Leon Kingsley

21305101

Euclid Kosasih

21315919

Khanh Tan (Jamie) Phan

21326604

Luke Warner

20930781

Table of Contents

List of Figures

List of Tables

List of Acronyms

FFT	Fast Fourier transform
LS	Least squares
LSE	Least squares error
MMSE	Minimum mean square error
SOE	Signal operating environment
SDA	Steepest descent algorithm

1 Introduction

1.1 Background

An on-going investigation has sampled a recording between two government ministers discussing a customer survey company. This recording is pivotal to the investigation. Unfortunately, the recording of the conversation is unintelligible due to background noise.

It is identified that two key sources of noise exist; $v_1[n]$ the noise caused by a background theme tune and $v_2[n]$ the noise cause by background crowd chatter. In addition to these noise sources a hum $v_3[n]$, caused by equipment issues, is suspected to be also corrupting the conversation.

Further investigation also reveals that the acoustic environment of where the recording was sampled was designed to provide an RT60 reverberation time of between 20ms and 30ms.

1.2 Aims

The objective of this project is to:

- Provide a method to estimate the desired signal $y[n]$,
- To ultimately, determine the information conveyed during the conversation from the estimate $\hat{y}[n]$.

Given:

- A noisy recording $x[n]$ of the conversation,
- Knowledge of the fact that the recording is corrupted by three distinct noise sources, with reference recordings of two of the noise sources $v_1[n]$ and $v_2[n]$, and knowledge of the fact that the third noise source $v_3[n]$ is a hum across the entire recording.
- The spectrogram of the desired signal $y[n]$, and
- Knowledge of the fact that acoustic environment of the recording was designed to provide an RT60 reverberation time of between 20 ms and 30 ms.

2 Design Methodology

The team first analysed the material facts of the problem, as well as the assumptions needed to tackle the problem; these are summarised in section 2.1 (Facts, Observations & Assumptions). With these, a theoretical framework which served as the basis of the design is outlined in section 2.2 (Theoretical Framework).

Despite having a theoretical framework and a high-level architecture, several design solutions were possible. Each case was considered and implemented as described in section 2.3 (Design Architecture). To elect the best design solution, performance measures described in section 2.4 (Design Performance Measurement) were used.

2.1 Facts, Observations & Assumptions

In determining the best design method for this case, several facts and assumptions are outlined in Table 1.

Table 1: Summary of key developments and assumptions in consideration for design solution.

No	Development/Assumption	Source	Design Consequence
1	This is an unsupervised problem with no known reference of the desired signal, $y[n]$.	As stated in section 1.2 (Aims)	Cannot use optimum filtering as we do not have *a priori* knowledge of the desired output.
2	There are known references to the noise sources $v_1[n]$ and $v_2[n]$.	As stated in section 1.2 (Aims)	We can construct an approximate signal model to estimate the correlated noise sources present in the noisy signal $x[n]$
3	The noise references presented produce a wide range of frequencies in a non-regular pattern.	Observation made from noise references.	It would be ineffective in only using filters which attenuate frequencies at a particular cut-off/band.
4	The noise sources that corrupts the measurement $x[n]$ are uncorrelated to the desired signal $y[n]$ but are correlated in some manner to their respective references.	Assumption	Estimates of the noise sources can be made from the noise references and subtracted from the measurement to provide an estimate of the desired signal
5	The corrupting noise sources themselves are not correlated to one another in the measurement $x[n]$	Assumption	The noise sources can be individually treated in designing the filter.
6	The SOE is such that the RT60 time is approximately 20 – 30ms, assumed to be for all frequencies, and the SOE is assumed to be constant.	Assumption	Block adaptive filters can be used without needing to re-estimate properties
7	The hum produced by $v_3(n)$ is produced at a constant, limited band of frequencies across the whole recording but is not subject to the SOE.	Assumption	A band-stop filter is required to attenuate this hum frequency.

2.2 Theoretical Framework

The problem at hand is to design a filter in which can estimate $y(n)$ from $x(n)$, where:

$$x(n) = y(n) + h_1(n) * v_1(n) + h_2(n) * v_2(n) + v_3(n)$$

With h being an unknown acoustic impulse response functions that map the known reference signals to the observed signal $x[n]$, representing the SOE.

From item 4 in Table 1, we assume that the corrupting signals $h_1(n) * v_1(n)$, $h_2(n) * v_2(n)$ and $v_3(n)$ are statistically independent to the desired signal $y(n)$, and further that the transformed corrupting signals are correlated in some manner to their respective references $v_1(n)$ and $v_2(n)$.

We also assume the pure tone or 'hum' $v_3[n]$ has a constant bandwidth and is not transformed by h, as it is stated to have been due to external factors (item 7 in Table 1), and hence:

$$v_3(n) = cos(\Omega n + \Phi)$$

Now, without *a priori* knowledge of $y(n)$ it is impractical to use exact filters that rely on optimum filtering. However, with references to the corrupting noises as described above we are able to produce estimates, such that:

$$\hat{v}_1(n) \approx h_1(n) * v_1(n)$$

$$\hat{v}_2(n) \approx h_2(n) * v_2(n)$$

$$\hat{v}_3(n) \approx v_3(n)$$

Assuming the individual noise sources are not correlated to one another (item 5 in Table 1):

$$x(n) \approx y(n) + \hat{v}_1(n) + \hat{v}_2(n) + \hat{v}_3(n)$$

And hence:

$$y(n) \approx \hat{y}(n) = y(n) + \left(h_1(n) * v_1(n) - \hat{v}_1(n)\right) + \left(h_2(n) * v_2(n) - \hat{v}_2(n)\right) + \left(v_3(n) - \hat{v}_3(n)\right)$$

$$= x(n) - (\hat{v}_1(n) + \hat{v}_2(n) + \hat{v}_3(n))$$
$$= x(n) - \hat{y}_c(n) - \hat{y}_3(n)$$

Where:

$$\hat{y}_c(n) \triangleq x(n) - (\hat{v}_1(n) + \hat{v}_2(n))$$

The problem has thus become a matter of estimating the respective noise sources with minimal error and filtering these estimates from the noisy signal $x(n)$. In particular, as a consequence of assumptions 4, 5 & 7 in Table 1, we identify two distinct stages of filtering which must occur to implement the design solution as shown in Figure 1.

We have the two filters H and B, with H responsible for cleaning the signal from the noise sources $h_1(n) * v_1(n)$ and $h_2(n) * v_2(n)$ provided their respective references and B the band-stop filter to remove the hum.

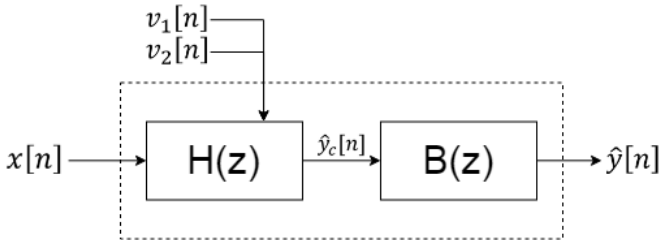

Figure 1: High level block diagram of the filtering stages.

2.3 Design Architecture

As described in section 2.2, the general solution requires two filtering stages $H(z)$ and $B(z)$. We further expand on this by recognising that $H(z)$ comprises of two distinct sub-systems $H_1(z)$ and $H_2(z)$ – the filters for the individual noise sources $v_1(n)$ and $v_2(n)$, separable by reason of assumption 5 in Table 1.

The general design architecture is illustrated in Figure 2.

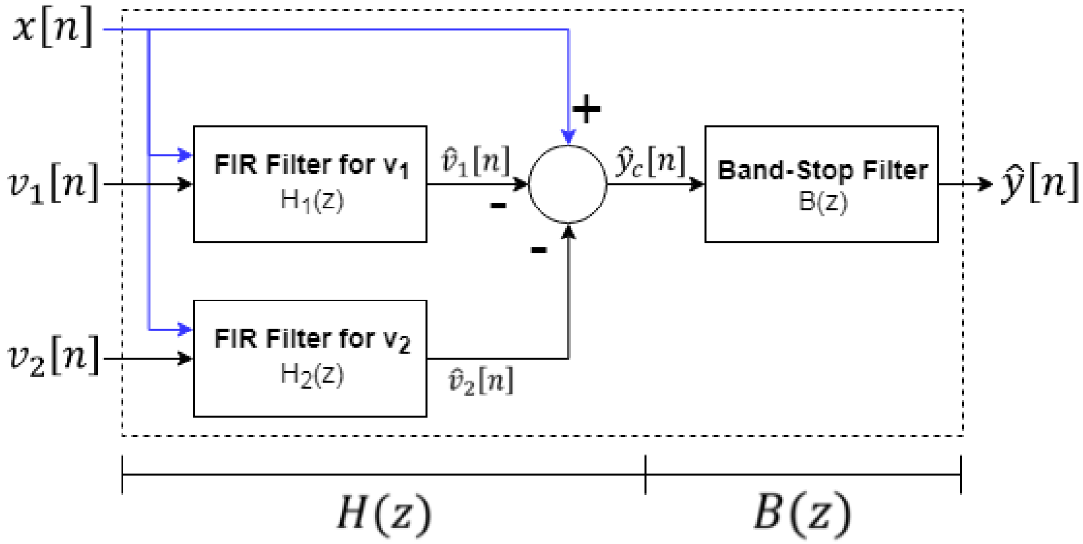

Figure 2: Block diagram of general architecture.

2.3.1 Band-Stop Filtering

A band-stop filter attenuates frequencies within a specified band – effectively 'rejecting' these frequencies whilst permitting other frequencies to pass through.

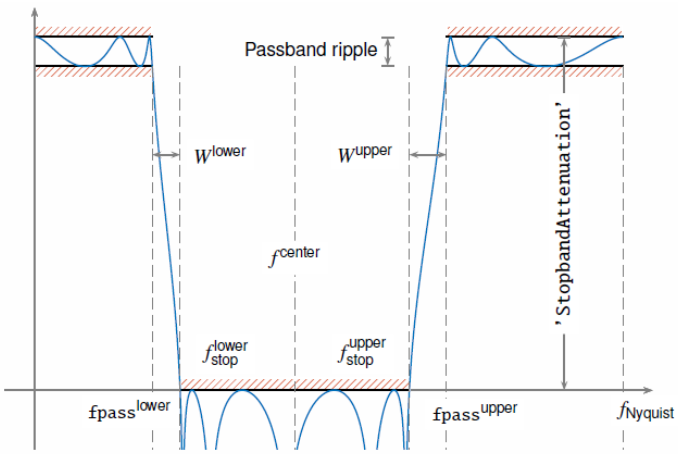

Figure 3: Illustration of band-stop filter in the frequency domain [2].

2.3.2 Least Squares Filter

Similar to optimum MMSE estimation, the LS estimation provides an estimate of a desired signal $d(n)$, with N samples over an interval $n \in [0, N-1]$, using a linear combination of M input signals, known as the *training data*, $x_k(n)$, $k \in [1, M]$. Formally:

$$\hat{d}(n) = \sum_{k=1}^{M} c_k(n) \cdot x_k(n)$$

Where $\hat{d}(n)$ is the estimate of the desired signal, $x_k(n)$ is the k^{th} data observation, and $c_k(n)$ is the scalar coefficient for the k^{th} data observation.

We thus define the estimation error $e(n)$:

$$e(n) \triangleq d(n) - \hat{d}(n)$$

In which the LS filter creates an estimate by minimizing the sum of squared errors, E:

$$E = \sum_{n=0}^{N-1} |e(n)|^2$$

Evidently, the LS filtering method is a deterministic method, in which the LSE estimator is dependant on the training data and the LS filtering problem becomes a matter of determining the filter coefficients $\boldsymbol{c}$ which minimise the sum of squared errors E. This can be determined by interpreting the LS estimate geometrically. We express the estimate in matrix form, with the $N \times 1$ column vectors $\boldsymbol{D}$, $\widehat{\boldsymbol{D}}$, $\boldsymbol{C}$, and $\boldsymbol{e}$ as well as the $N \times M$ training data matrix $\boldsymbol{X}$:

$$\widehat{\boldsymbol{D}} = \boldsymbol{C}^T \cdot \boldsymbol{X}$$

$$\boldsymbol{e} = \boldsymbol{D} - \widehat{\boldsymbol{D}}$$

Since the estimate is a linear combination of the M observations from the training data, it follows that the estimate $\widehat{\boldsymbol{D}}$ must lie within the subspace, $S = \text{span}(\boldsymbol{X})$, spanned by these observations (the column vectors in $\boldsymbol{X} = [x_1(n) \quad \dots \quad x_M(n)]$):

In contrast, for $\boldsymbol{e} \neq \emptyset$ the desired response $\boldsymbol{D}$ does not lie within S. In such cases, the minimum error, let us denote with $\boldsymbol{e_{ls}}$, is such that the vector $\widehat{\boldsymbol{D}}$ is a direct projection of $\boldsymbol{D}$ onto the S, or in other words the error vector $\boldsymbol{e_{ls}}$ is orthogonal to S.

$$\langle x_k, e_{ls} \rangle = x_k^T \cdot \boldsymbol{e} = 0 \quad \forall\, k \in [1, M]$$

$$= \boldsymbol{X^T e} = \boldsymbol{X^T}(\boldsymbol{D} - \boldsymbol{X c_{ls}})$$

$$\Rightarrow (\boldsymbol{X^T X})\boldsymbol{c_{ls}} = \boldsymbol{X^T D}$$

$$\widehat{\boldsymbol{R}}\boldsymbol{c_{ls}} = \boldsymbol{d}$$

2.3.3 Recursive Least Squares Filter

For the RLS filter, a weighted MSE is used [3]:

$$E(n) = \sum_{j=0}^{n} \lambda^{n-j} |e(j)|^2 = \sum_{j=0}^{n} \lambda^{n-j} |y(j) - \boldsymbol{c}^H \boldsymbol{x}(j)|^2$$

This is the weighted sum of the total squared error since time $j = 0$, where λ is the forgetting factor, it applies an exponentially increasing weight to samples closer in time. The filter coefficients that minimise the total squared error are specified by the normal equations:

$$\widehat{\boldsymbol{R}}(n)\boldsymbol{c}(n) = \widehat{\boldsymbol{d}}(n)$$

The design of the forgetting factor is dependent on the SOE. With a forgetting factor of 1, the RLS filter will consider each data point with the same weighting, and will thus, be unable to pay less attention to past data, making it only suitable in stationary SOEs. To make it suitable for use in non-stationary SOEs, decreasing the forgetting factor will allow the RLS algorithm to place less weighting on past data, and be able to track changes in that environment.

2.3.4 Least Mean Square Filter

The LMS adaptive filter is a form of stochastic steepest-descent algorithm that minimises the instantaneous MSE. The optimal filter coefficients $\boldsymbol{c}_o(n)$ are determined by following the negative gradient of the MSE cost function $\zeta(n)$ which is approximated by the instantaneous MSE

$$\zeta(n) = E\{|e(n)|^2\} \approx |e(n)|^2$$

The SDA learning rate is controlled by the step size parameter μ when updating the coefficients

$$\boldsymbol{c}(n) = \boldsymbol{c}(n-1) + 2\mu \boldsymbol{x}(n) e(n)$$

2.4 Design Performance Measurement

Measuring the relative performance of various filter implementations required four key parameters to be analysed.

2.4.1 Intelligibility

Each design solution was to output a recording of the enhanced signal, and the team would listen to the output. Whether or not an intelligible conversation could be heard was an important measurement for determining the viability of the design solution. Due to the subjective nature of this, and inability to discriminate outputs which have slight noise, this measurement was only used to discriminate between the clearly incorrect and potentially correct.

2.4.2 Spectrogram

Given reference to the spectrogram of the original recording, each solution's output was plotted in a spectrogram and visually compared to the benchmark spectrogram. This is necessary to ensure that intelligible conversations produced by design solutions aligned with the desired conversation. Due to the lack of digitization of the benchmark, visual comparison 'by eye' was performed, and consequently this measurement was only used to discriminate between the clearly incorrect and potentially correct.

2.4.3 Signal to Noise Ratio

As the previously mentioned qualitative methods depended on an individual's own interpretation of performance, which may be prone to error, a quantitative performance measurement was needed. The signal to noise ratios between the noisy signal $x(n)$ and noises v_1 and v_2 were found. This was set as a reference point in which the signal to noise ratios between the filtered signals and noises v_1 and v_2 were compared to. It is expected that as a filter removes more noise component, the signal to noise ratio between the filtered signal and that noise component will increase.

2.4.4 Computational Complexity

An important consideration was the computational resources required to provide an output. This measurement consisted of measuring the *wall-clock* time of each design solution under controlled computational environments. CPU time was not measured due to the inability to restrict the level of threading and consequently the variability of the CPU time measurement was considered inaccurate.

3 Experimental Evaluations

3.1 Summary of Design Iterations

Table 2 shows the various iterations employed to design the filter. It will be shown in Section 0 that the band-stop filter removes the hum v_3 almost perfectly and so it was used for all iterations 1-5. The band-stop filter worked very well for removing noise which occurred at very specific frequencies however for the more complex noise signals v_1 and v_2, adaptive filters had to be used. For v_1 and v_2, the iterations involved experimenting with different combinations of NLMS, RLS and LSE filters, as shown in Table 2.

Table 2: Design iterations.

Iteration	**Filter**		
	v_1	v_2	v_3
1	NLMS	NLMS	Band-stop
2	RLS	LMS	Band-stop
3	LMS	RLS	Band-stop
4	RLS	RLS	Band-stop
5	LSE	LSE	Band-stop

The spectrogram of the noisy observation can be seen below in Figure 4. By comparing it to the spectrogram of the desired signal in Figure 5, it can be seen that the noisy signal consisted of many horizontal lines that were superimposed onto the clean desired signal, represented by vertical lines. This is not surprising as the crowd noise resembles an almost drone like noise that drowned many different frequencies.

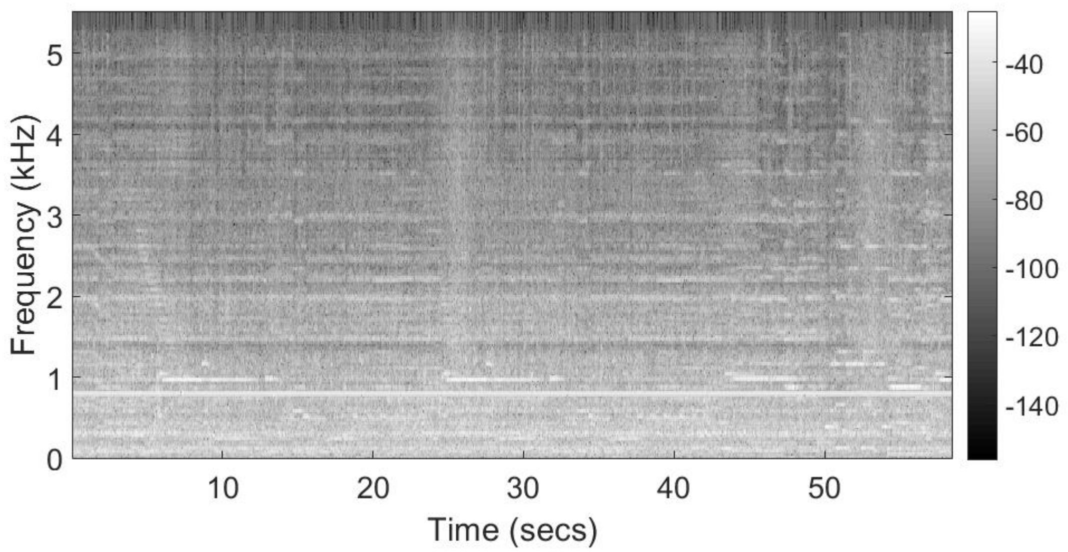

Figure 4: Spectrogram of noisy observation, $x(n)$

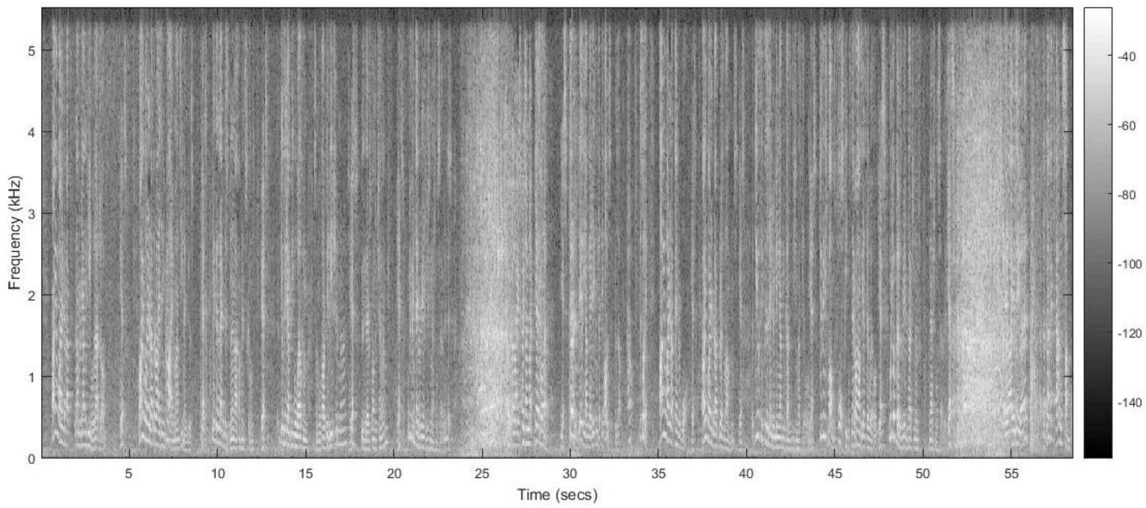

Figure 5: Spectrogram of desired signal, $y(n)$

The signal to noise ratio between the noisy observation and v_1 and v_2 were –2.7582 dB and 1.5449 dB respectively.

3.2 Band-Stop Filter

The spectrogram analysis of the given signal (Figure 4) revealed a line throughout the entire recording, just below the 1 kHz line. Taking the FFT of the signal shows that the line is a very specific frequency, 799.995 Hz.

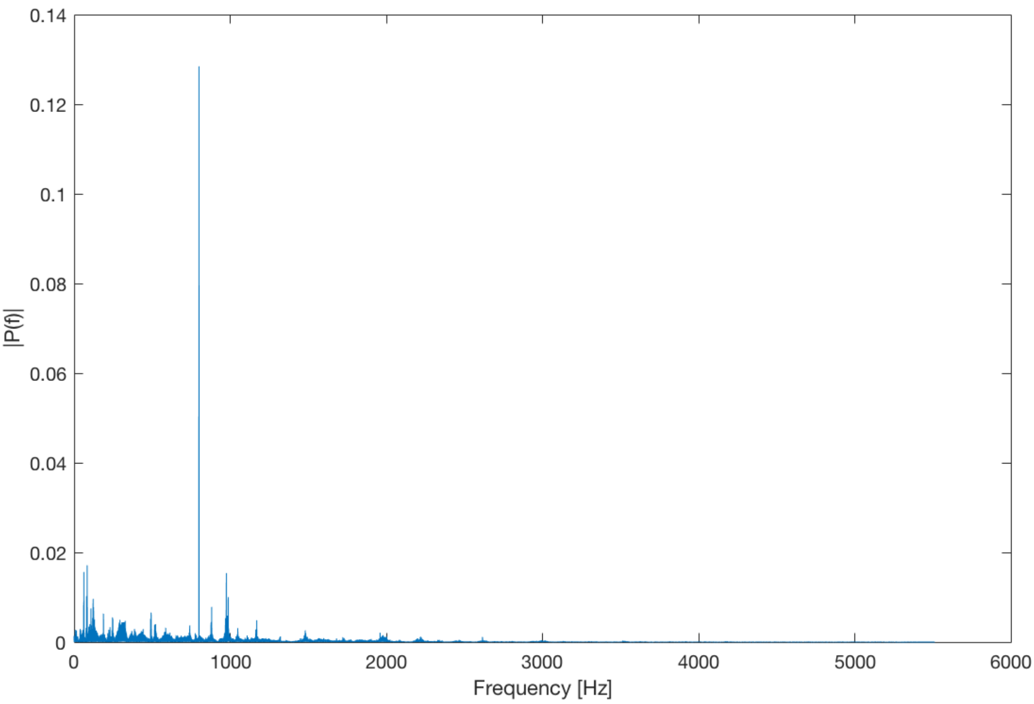

Figure 6: FFT of Noisy Signal

The single frequency signal or pure tone can be removed by means of a Band-stop filter. The band-stop filter was implemented by using the built-in function ”`bandstop`”. The upper and lower frequencies were chosen by trial and error, running the signal through the filter, doing a

FFT to judge the performance until the 'hum' was removed without removing too much of the surrounding frequencies. We found that the ideal range for the "`bandstop`" filter to be between 790 and 810 Hz.

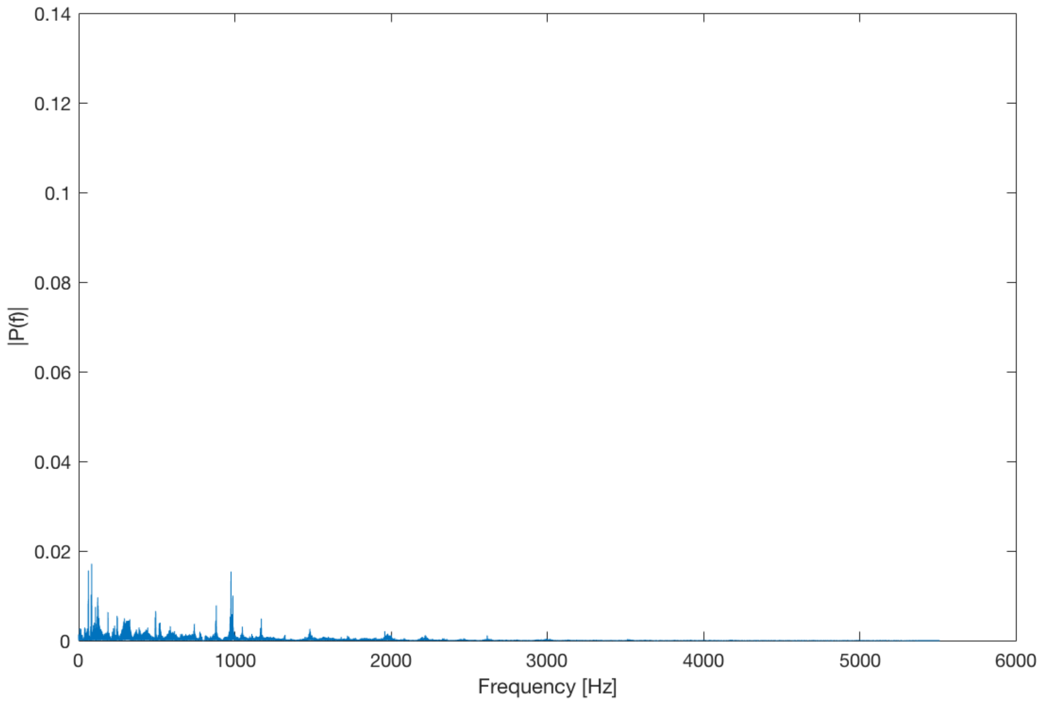

Figure 7: FFT of Noisy Signal after Band-stop Filtering

The band-stop filter worked very well on removing specific frequencies from the recording, but due to the static nature of the filter, it cannot be used to remove complex signals like music or speech. The band-stop filter was used in all design iterations to remove the pure tone v_3.

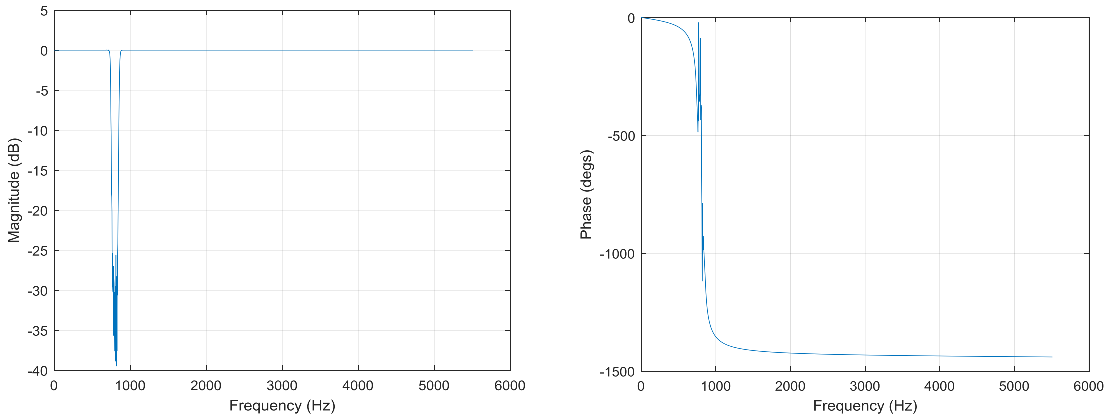

Figure 8: Magnitude (left) and phase (right) response of the Butterworth band-stop filter used to remove the hum.

3.3 Filter Order

The filter order for was determined from the SOE; an RT60 reverberation time, T_{60}, of 20 – 30ms. The reverberation created in the SOE will produce delayed versions of all the audio signals present, in which will only decay to an insignificant level after a time T_{60} has elapsed. Clearly then, this suggests that in creating FIR filters which estimates signals from measurements in this SOE, the training data used to train the filter must encompass all of these delayed signals.

That is to say, reverberation will continue up to T_{60}, permitting a maximum delay $x(n - N_{60})$ where $N_{60} \in \mathbb{Z}^+$ is the corresponding index for $t = T_{60}$; this can be determined with the sampling frequency, f_s

$$N_{60} = \text{round}\{f_s \cdot T_{60}\}$$

Hence the best-estimate, $\hat{y}(n)$ must exist in the subspace spanned by the vectors, $\{x(n), x(n-1), \dots, x(n - N_{60})\}$; and hence the filter order must be at least N_{60}.

3.4 NLMS (v_1), NLMS (v_2) & Band-Stop Filter (v_3)

The application of NLMS adaptive filters on both the Dr Who theme and crowd noise provided the first insight into the hidden speech. Although both noises can be heard, their volumes were much lower, as if they were much further away. This allowed the full speech to be deciphered. An adaptation gain of 0.05 was used

3.4.1 Spectrogram Comparison

Figure 9 depicts the spectrogram of the filtered signal. When compared to the spectrogram of the noisy signal in Figure 4, it can be clearly seen that there is much less noise in the filtered signal. Vertical lines representing the hidden speech signal can finally be seen across the spectrogram, with less horizontal lines running across the higher frequencies.

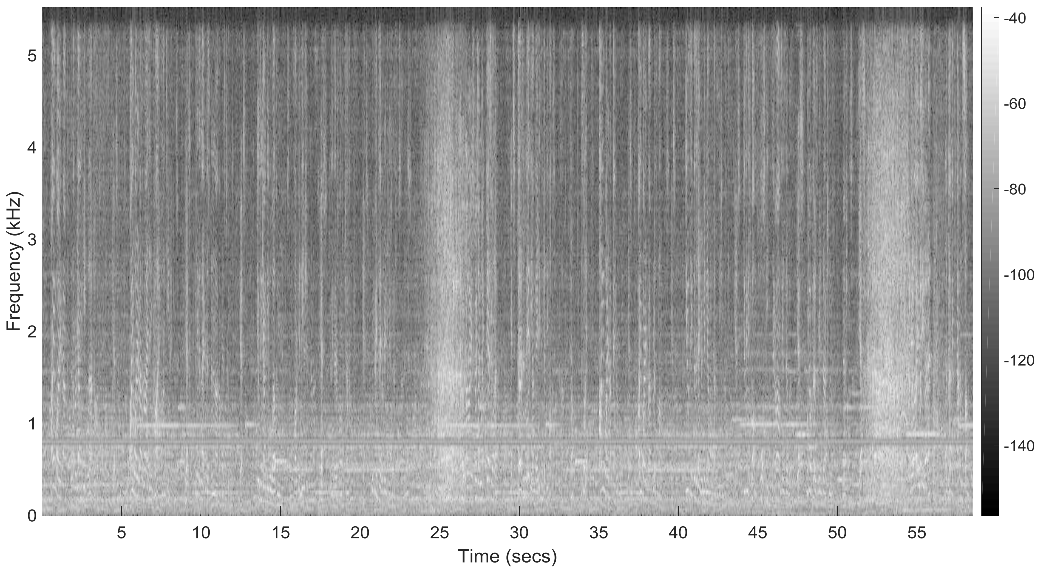

Figure 9: Spectrogram of the cleaned signal after applying NLMS filters for noises v_1 and v_2 and a band-stop filter for noise v_3.

3.4.2 Signal to Noise Ratio

The signal to noise ratios between the filtered signal and v_1 and v_2 were -6.9910 dB and -2.6879 dB respectively, much larger than the signal to noise ratios of the original noisy signal introduced in section 3.1. This marks a significant improvement to the noisy signal.

3.4.3 Computational Complexity

Simulation of the NLMS adaptive filter implementation found that it had an average wall-clock time of 1.9267 s.

3.4.4 Parameter Tuning

From Figure 10 to Figure 15, it is apparent that decreasing the step size μ from 1 to 0.01 improves the estimation (spectrogram more closely resembles the clean signal). However, past $\mu = 0.01$, decreasing the step size makes the output noisy again. This is because the filter order is kept constant at $N = 100$. For too small of a step size ($\mu < 0.01$), $N = 100$ is not enough iterations to allow the filter coefficients to converge. For an accurate solution using an NLMS filter, we simultaneously need a sufficiently small step size and a sufficiently large filter order to allow time for the algorithm to convergence. We thus used a small a step size of $\mu = 0.001$ and increased the order to 250.

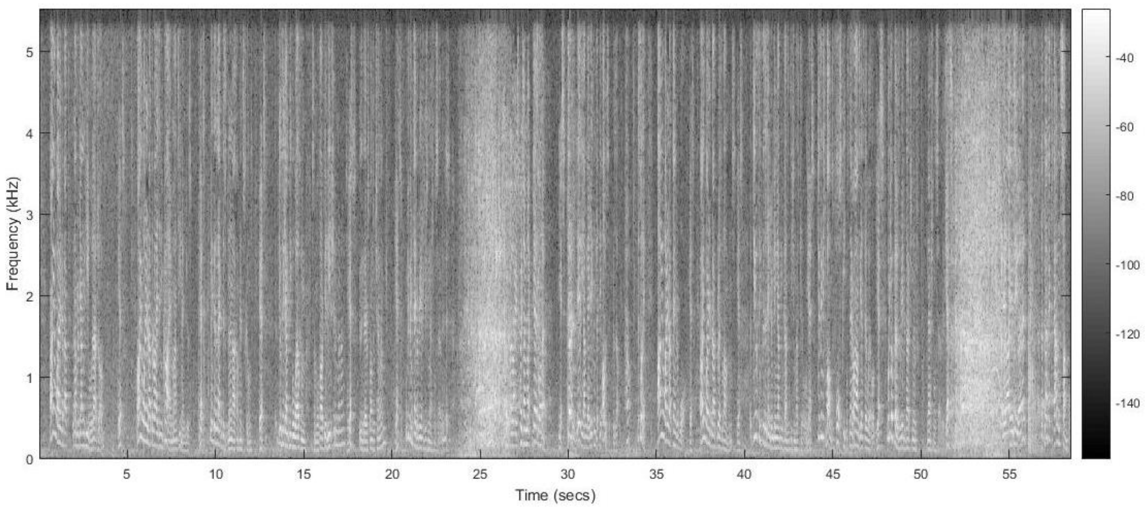

Figure 10: Spectrogram of desired signal.

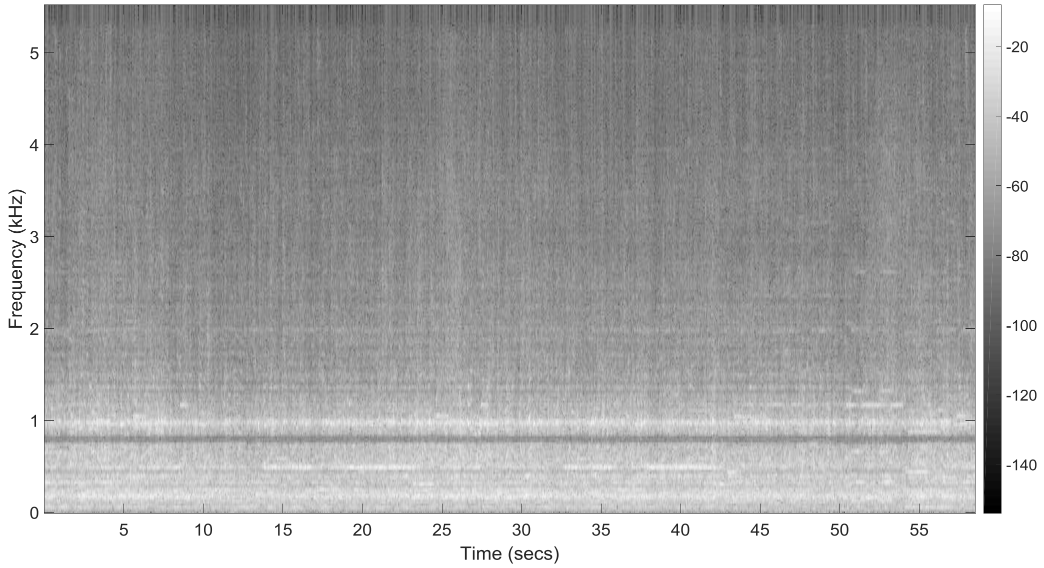

Figure 11: Spectrogram for $\mu = 1$ and $N = 100$

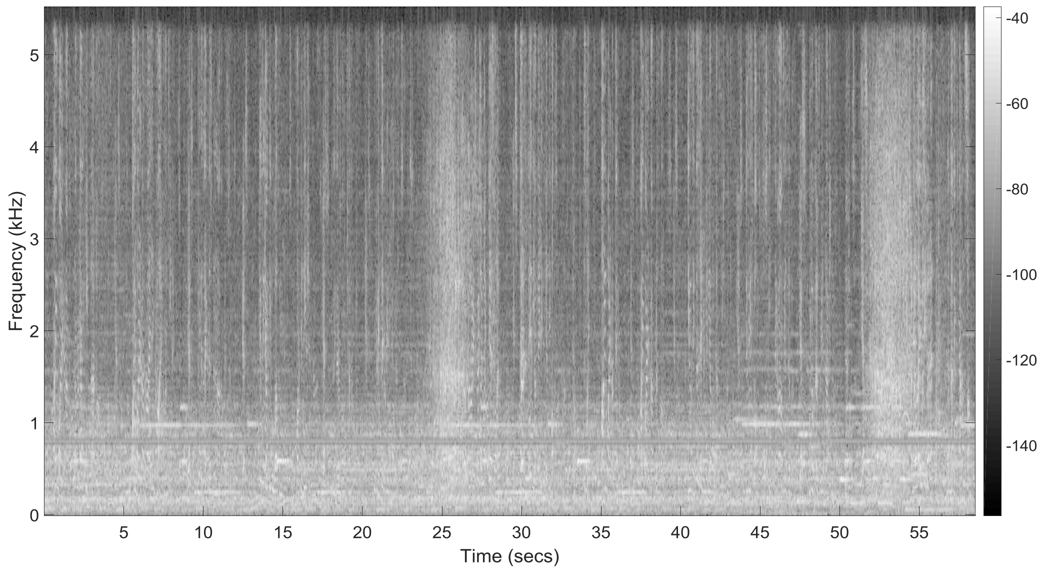

Figure 12: Spectrogram for $\mu = 0.1$ and $N = 100$

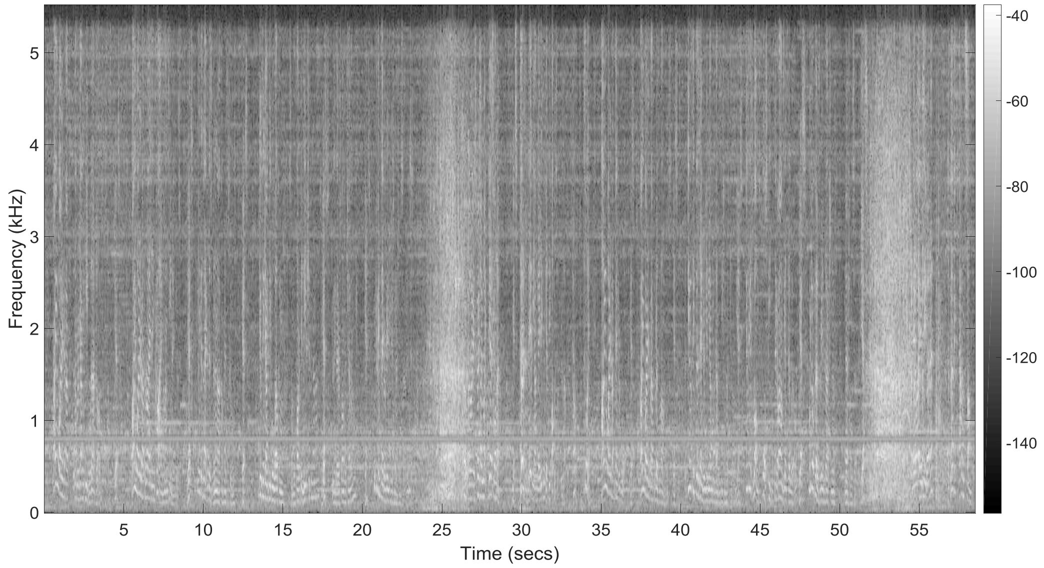

Figure 13: Spectrogram for $\mu = 0.01$ *and* $N = 100$

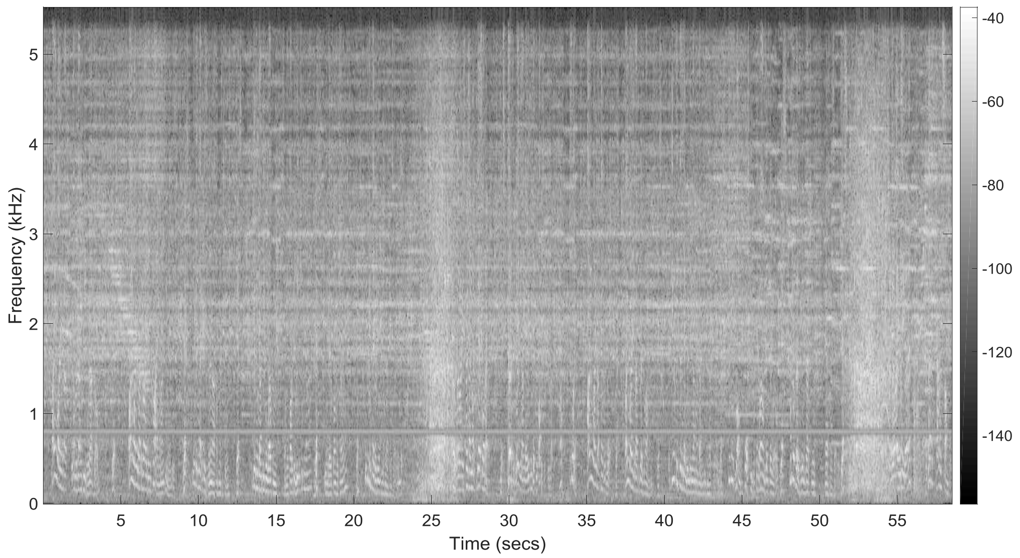

Figure 14: Spectrogram for $\mu = 0.001$ *and* $N = 100.$

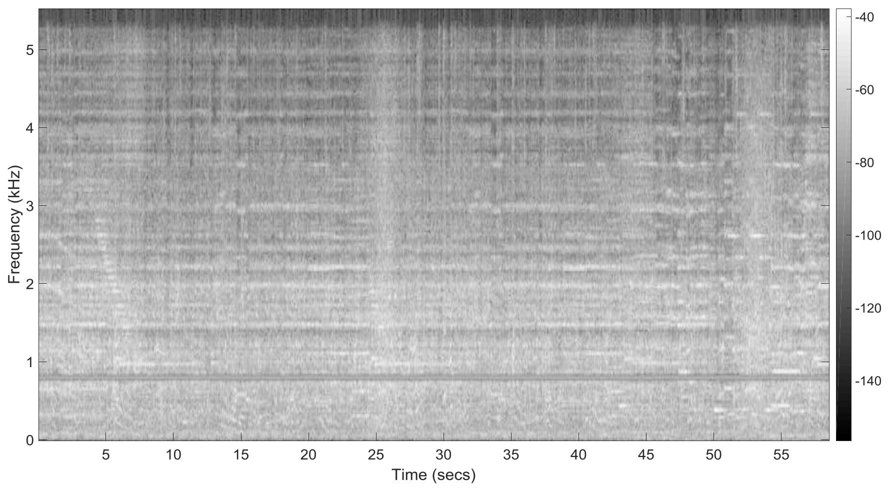

Figure 15: Spectrogram for $\mu = 0.0001$ and $N = 100$.

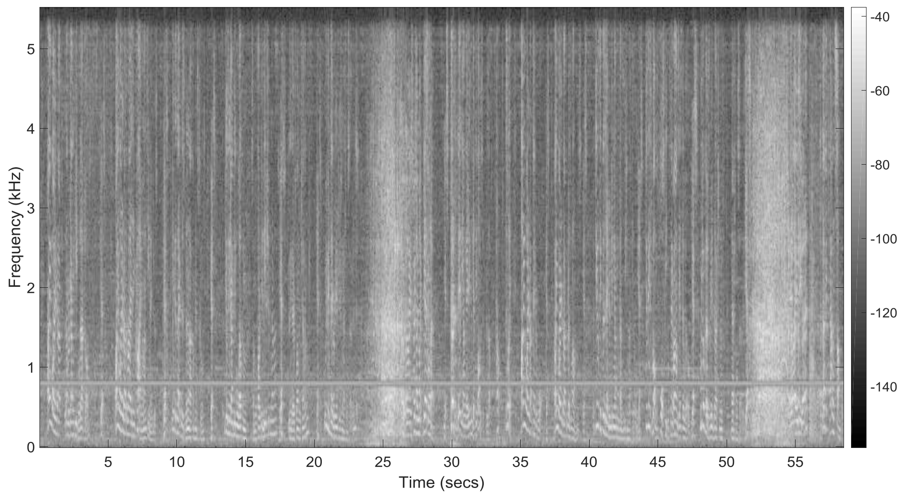

Figure 16: Spectrogram for NLMS filter with parameters $\mu = 0.001$, $N = 250$.

3.5 RLS (v_1), NLMS (v_2) & Band-stop Filter (v_3)

The use of an RLS filter on the Dr Who theme instead of the NLMS used previously produced an interesting result. Listening to the output audio file, although the crowd noise can still be heard, the Dr Who theme was almost non-existent. This suggests that the RLS filter may produce the best result. The filtered signals produced by the RLS filter implementation did not work well with a forgetting factor set any less than $\lambda = 1$. This does not come as a surprise as mentioned in assumption 6, in a stationary SOE, the RLS algorithm does not need to track changes to the environment. As such, all previous observations need not be 'forgotten' and will provide more accurate data.

3.5.1 Spectrogram Comparison

Looking at the spectrogram in Figure 17, a minor improvement can be seen to the previous filter design with less noise in the higher frequencies.

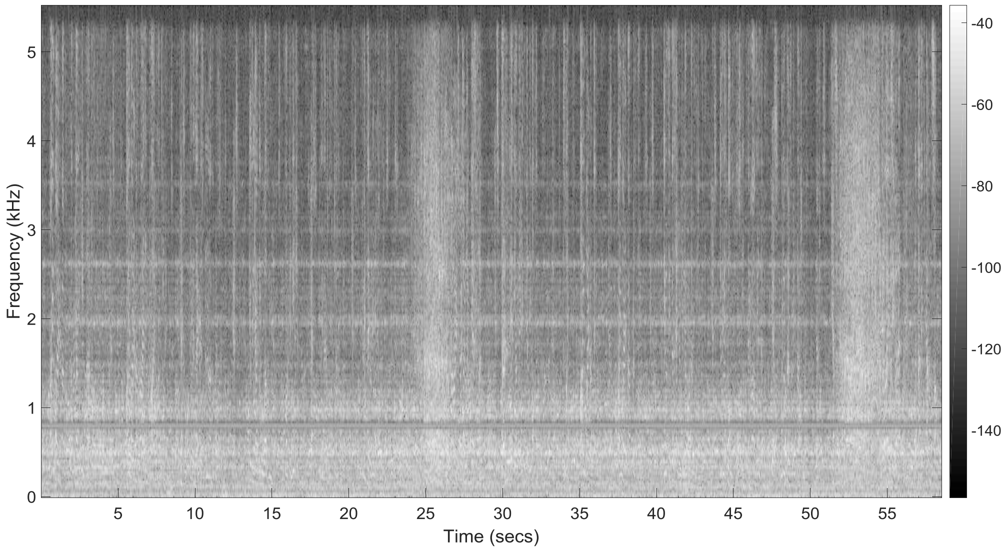

Figure 17: Spectrogram of the enhanced signal after applying RLS to v_1, NLMS to v_2 and band-stop filter to v_3.

3.5.2 Signal to Noise Ratio

An improvement can be seen in the signal to noise ratio with it being -7.2463 dB and -2.9432 dB between the filtered signal and noises v_1 and v_2 respectively.

3.5.3 Computational Complexity

The addition of an RLS filter greatly increased the wall-clock to around 125.0199 s. This is expected due to the increased complexity of the RLS filter.

3.6 NLMS (v_1), RLS (v_2) & Band-stop Filter (v_3)

Applying an LMS filter to the Dr Who theme and an RLS filter to the crowd noise further supported the previous suspicion that the RLS filter may produce the best results. It could be heard in the audio recording that although the speech can be deciphered, the Dr Who theme was still quite prominent. The crowd noise however, was now very well suppressed.

3.6.1 Spectrogram Comparison

The spectrogram in Figure 18 showed more prominent vertical lines when compared to previous filter designs. Less noise can be seen throughout the spectrogram with the horizontal lines produced by the crowd noise visibly reduced.

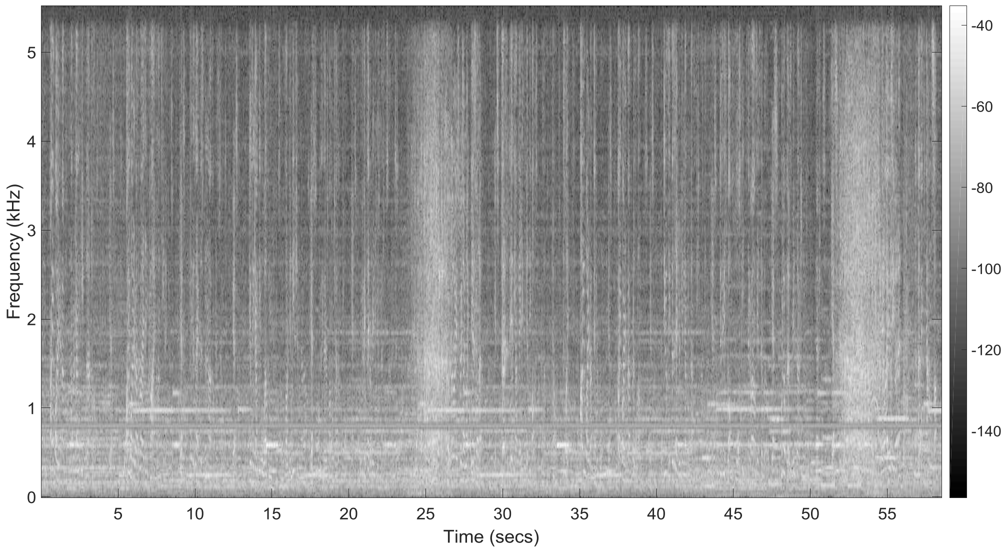

Figure 18: Spectrogram of the enhanced signal after applying NLMS filter for v_1, RLS filter for v_2 and band-stop filter for v_3.

3.6.2 Signal to Noise Ratio

The signal to noise ratios to v_1 and v_2 were -7.3619 dB and -3.0588 dB respectively, a small improvement to the previous filter design.

3.6.3 Computational Complexity

As seen earlier, the RLS filter introduces a much greater wall-clock time, with this filter implementation having a running time of 160.9028 s. This was expected of the RLS algorithm with $O(4N^2)$, this algorithm is the most computationally complex of the selection.

3.7 RLS (v_1), RLS (v_2) & Band-Stop Filter (v_3)

Of all the filters, the RLS filter took the most resources to execute. With just a single iteration of the RLS algorithm on both noises, the runtime for the code was over 260 seconds. However, it was the best signal produced thus far, as both noises were suppressed very well, and the speech can be heard very clearly.

3.7.1 Spectrogram Comparison

Figure 19 depicts the spectrogram of the filtered signal where strong vertical lines can be seen running down the bottom. This is a significant improvement to all previous filter designs with the spectrogram almost resembling the desired signal.

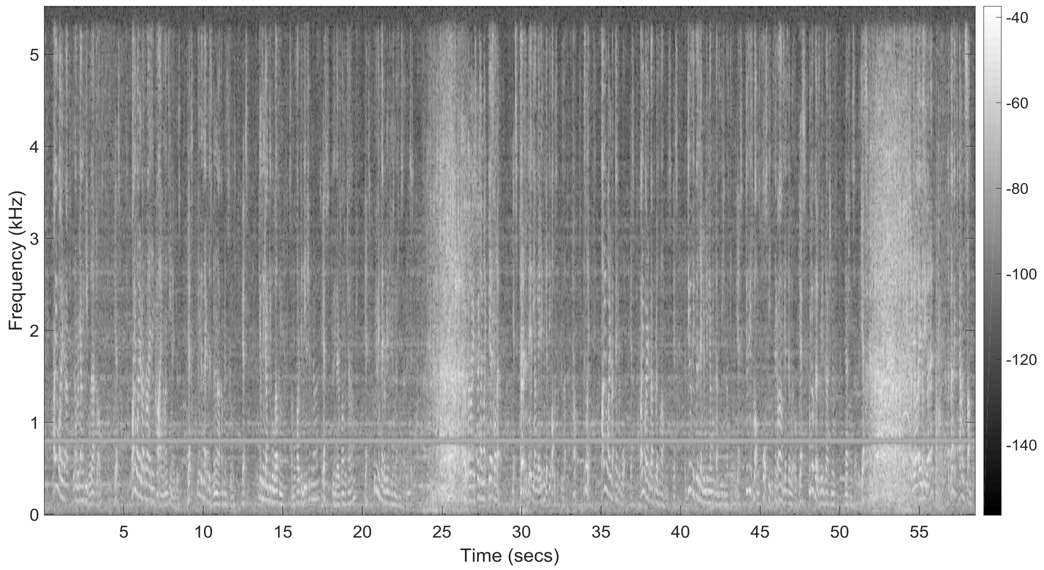

Figure 19: Spectrogram of the RLS filter and bandstop filter output ($\lambda = 1.0$)

3.7.2 Signal to Noise Ratio

An improvement was also seen with the signal to noise ratios being -7.6446 dB and -3.3415 dB compared to v_1 and v_2 respectively.

3.7.3 Computational Complexity

Although the complete RLS filter implementation produced good results, it had the longest wall-clock time of 266.7065 s. This was expected as the RLS algorithm is the most computationally complex algorithm amongst the selection, with $O(4N^2)$ [7]. This relatively high computational complexity, whilst provides greater filter stability, is a result of the RLS algorithm requiring large memory spaces to track previous estimates of the signal.

3.8 LS (v_1), LS (v_2) & Band-Stop Filter (v_3)

3.8.1 Spectrogram Comparison

Figure 20 shows a higher contrast between the vertical lines in contrast to the RLS variant reflecting improved speech clarity. This is demonstrated at approximately $f \in [2700\ Hz, 3300\ Hz]$ where a dark band can be observed in the LS spectrogram and not as evident in the RLS spectrogram. The 3kHz band is roughly the upper limit to the *usable voice frequency* and the higher contrast between the so called 'vertical bars' in the LS variants represents the filters capability in discriminating noise from speech – as desired.

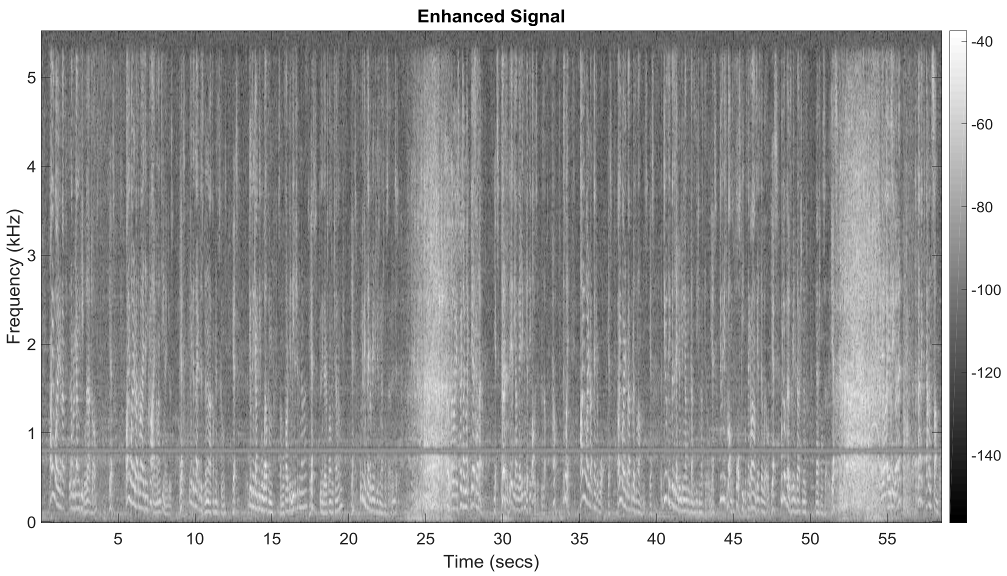

Figure 20: Spectrogram of enhanced signal outputted from LS filter

3.8.2 Signal to Noise Ratio

An improvement was also seen with the LS filter having a signal to noise ratio of -7.6487 dB and -3.3455 dB compared to v_1 and v_2 respectively.

3.8.3 Computational Complexity

A wall-clock time of only 8.056s was observed for the LS filter design. This is expected as the LSE estimator only requires calculation of the error signal on top of the linear combinations of the training data (as do all the algorithms).

3.8.4 LS Filter for v_1

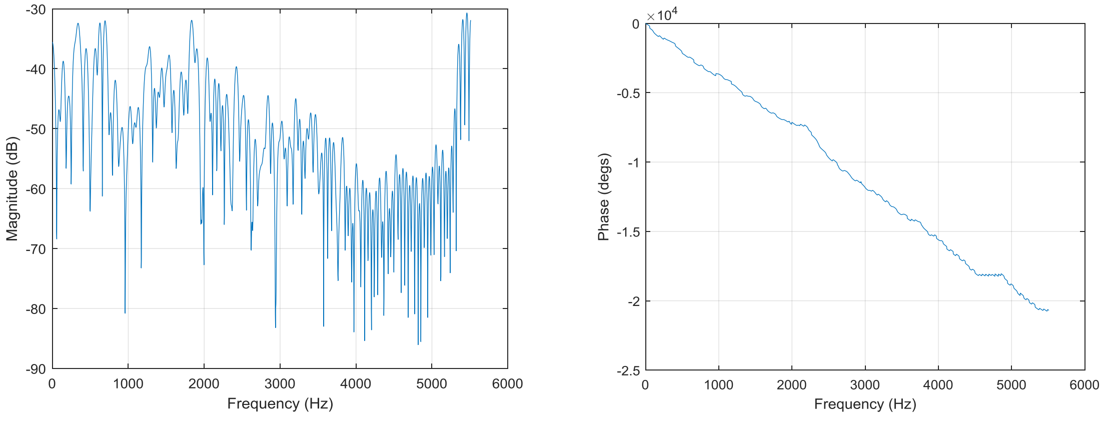

Figure 21: Magnitude (left) and phase (right) response of the LS filter applied to v_1.

3.8.5 LS Filter for v_2

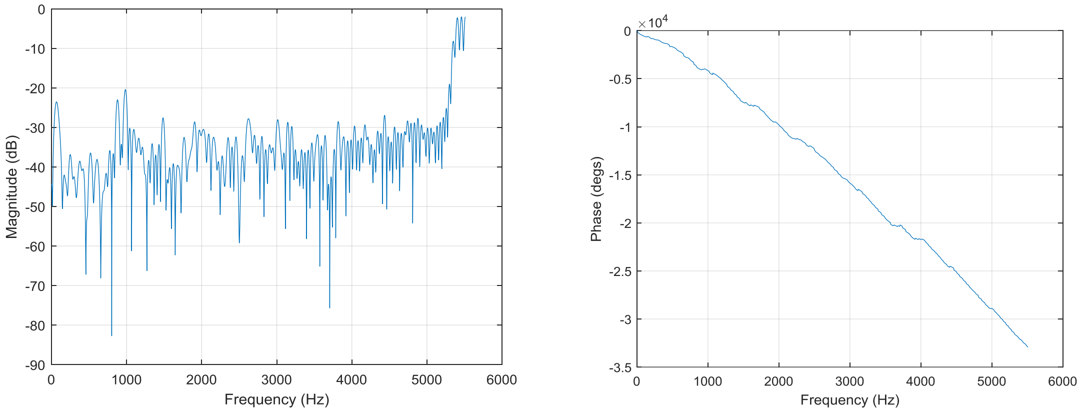

Figure 22: Magnitude (left) and phase (right) response of the LS filter applied to v_2.

Figure 21 and Figure 22 show that the phase response of the LS filter is linear, as expected.

3.9 Comparison

Table 3: Table comparing the relative performances of each filter design

	Performance Measure			
	Intelligibility	Spectrogram	Signal to Noise	Computational Complexity
NLMS	5th	5th	5th	1st
RLS(v_1)+NLMS(v_2)	4th	4th	4th	3rd
NLMS(v_1)+RLS(v_2)	3rd	3rd	3rd	4th
RLS	1st	1st	1st	5th
LS	1st	1st	1st	2nd

The differences between the filters can be seen best when comparing the spectrograms of each filter; in particular in examining the differential spectrogram between the reference signal v_1 and $\hat{v}_1$ as seen in Figure 23. It can be seen that at the $t = 14$ second mark (at the first audience laugh, see transcript in section 4.2) there are distinct differences between the NLMS filter compared to the other two RLS and LS filters. Here, the spectrograms show that the NLMS differential is smoother, suggesting that the NLMS is less capable in discriminating between frequencies and attenuates larger bands of frequencies together; in contrast the LS and RLS filters have higher resolution contours across the entire frequency band, $f \in [0{,}100\ Hz]$ shown.

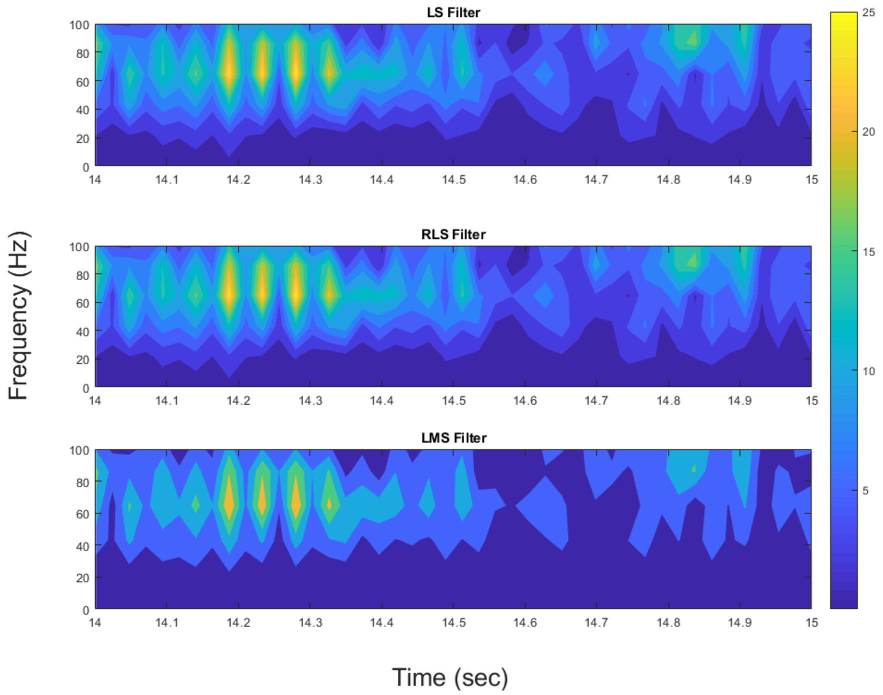

Figure 23: Comparison of differential $|v_1 - \hat{v}_1|$ spectrograms for each filter across $t \in [14{,}15\ s]$ and $f \in [0{,}100\ Hz]$

Indeed, the NLMS algorithm belongs to a family of stochastic gradient algorithms in which the update to the gradient is calculated from the instantaneous error signal, a noisy estimate of the gradient [6]. When recursively applied, this has the effect of effectively averaging the estimate [5]. Hence, for signals in which the frequencies appear random (as is the laughter), the noisy estimate may cause an averaging effect as observed. Similar behavior can be observed at $t = 25$ seconds (the second audience laughter).

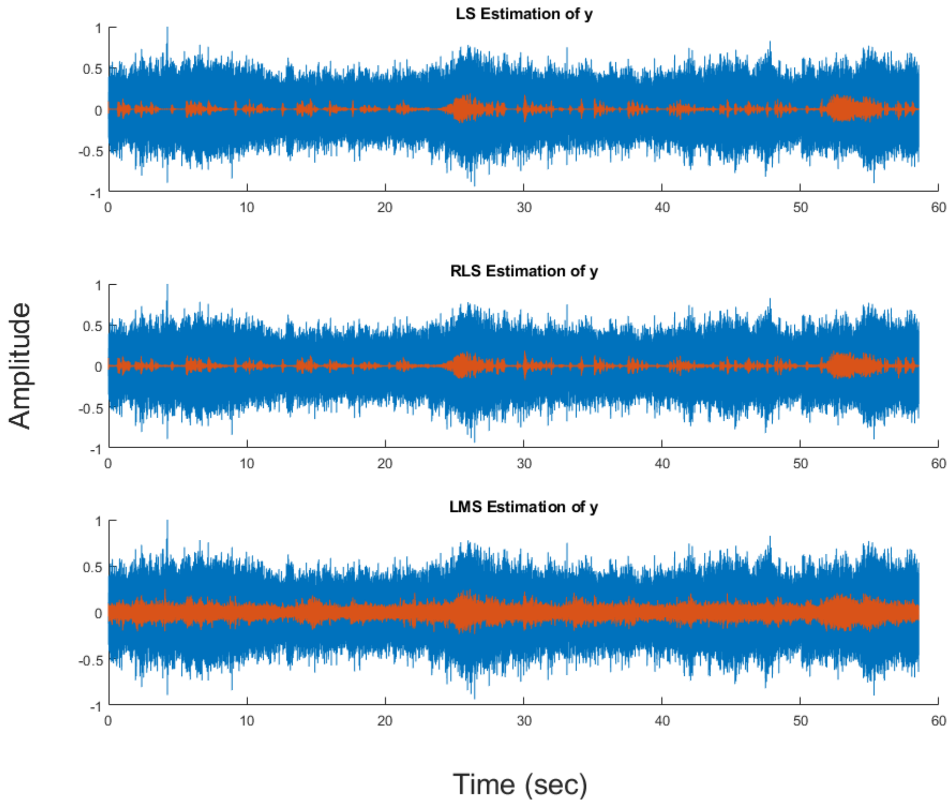

Figure 24: Comparison of the time series representations of the estimated signal from each of the filters

This averaging behavior is evident in the time-domain as seen in Figure 24. The estimate $\hat{y}_{LMS}$ produced by the NLMS filter consistently fluctuates near its peaks creating a thick 'band' across the entire signal, and consequently significantly noisier in comparison to the finer LS and RLS produced estimates.

4 Conclusion

The LS filter was chosen for its accuracy and quick computation time. When increasing the order of the filter and the number of iterations, there was a trade-off between the computation time and the accuracy of the enhanced signal. After tuning the parameters, the optimal values for the order filter and the number of loop iterations were 25 and 2 respectively. The final artefact, `EnhancedSignal.wav`, showed that using LS filters applied to both v_1 and v_2 and a band-stop filter to v_3 performed remarkably. The crowd noise, the Dr. Who theme and the hum were almost completely suppressed. The conversation could be heard almost perfectly and a transcript is produced in Section 4.2. The signal to noise ratio for the output signal using this method is:

RLS also had remarkable performance, however, the computation time was far longer, taking over 260 seconds for only one iteration on each of v_1 and v_2 and a forgetting factor of $\lambda = 1.0$.

The NLMS had mediocre performance as both the crowd noise and the Dr. Who theme were still audibly interfering with the conversation.

4.1 Limitations

The Butterworth band-stop filter was the easiest method of removing the hum, however, because of the band pass size, some of the desired signal was also attenuated. This lead to unwanted loss of power in the enhanced signal. This is most evidently seen in Figure 24, where the amplitudes of the filtered signal was significantly smaller to the original signal, and further can be seen in the spectrograms with reduced intensity at the $800\ Hz$ band.

Several assumptions were necessary to the design of this design solution, in particular items 4 and 5 in Table 1 which requires that the noise sources that corrupts the measurement $x[n]$ are uncorrelated to the desired signal $y[n]$ but are correlated in some manner to their respective references, whilst themselves being uncorrelated to one another. This is manifested in the theoretical framework which assumes the acoustic environment produces independent noise signals that can be simply subtracted such that:

$$\hat{y}_c(n) \triangleq x(n) - (\hat{v}_1(n) + \hat{v}_2(n))$$

With this assumption, the design is limited to cases in which the noise sources have noise references and are known to be uncorrelated.

4.2 Transcript

Two speakers were identified, one being named in the speech as ‘Mr. Wooley’.

[Speaker 1]:	*Are you worried about the number of young people without jobs?*
[Mr. Wooley]:	*Yes*
[Speaker 1]:	*Are you worried about the rise in crime among teenagers?*
[Mr. Wooley]:	*Yes*
[Speaker 1]:	*Do you think there’s a lack of discipline in our comprehensive schools?*
[Mr. Wooley]:	*Yes*
[Speaker 1]:	*Do you think young people welcome some authority and leadership in their lives?*
[Mr. Wooley]:	*Yes*
[Speaker 1]:	*Do you think they respond to a challenge?*
[Mr. Wooley]:	*Yes*
[Speaker 1]:	*Would you be in a favour of reintroducing National Service?*

[Crowd laughs]

[Mr. Wooley]:	*Oh well, I suppose I might.*
[Speaker 1]:	*Yes or no?*
[Mr. Wooley]:	*Yes*
[Speaker 1]:	*Alternatively, the young lady can get the opposite result.*
[Mr. Wooley]:	*How?*
[Speaker 1]:	*Mr. Wooley, are you worried about the danger of war?*
[Mr. Wooley]:	*Yes*
[Speaker 1]:	*Are you worried about the growth of armaments?*
[Mr. Wooley]:	*Yes*
[Speaker 1]:	*Do you think there's a danger in giving young people guns and teaching them how to kill?*
[Mr. Wooley]:	*Yes*
[Speaker 1]:	*Do you think it's wrong to force people to take up arms against their will?*
[Mr. Wooley]:	*Yes*
[Speaker 1]:	*Would you oppose the reintroduction of National Service?*
[Mr. Wooley]:	*Yes*

[Crowd laughs]

[Speaker 1]:	*There you are, you see Bernard. The perfect balanced sample.*

5 References

[1] Singh, A. “Adaptive Noise Cancellation”. Online: http://www.cs.cmu.edu/~aarti/pubs/ANC.pdf, May, 2001 [May 29, 2018].

[2] MathWorks. “Bandstop”. Online: https://au.mathworks.com/help/signal/ref/bandstop.html, 2018 [May 30, 2018]

[3] D.G. Manolakis, V.K. Ingle, S.M. Kogon, “Statistical and Adaptive Signal Processing”, McGraw-Hill, 2000. (Chapter 10)

[4] A.V. Oppenheim, A.S. Willsky, “Signals and Systems”, Prentice Hall, 1983.

[5] Kaur, H., Malhotra, R., and Patki, A., “Performance analysis of gradient adaptive LMS algorithm”. *International Journal of Scientific and Research Publications.* [Online]. 2(1), pp. 1-4. [Jun. 1, 2018].

[6] Jaggi, S. and Martinez, A.B. “Upper and lower bounds of the misadjustment in the LMS algorithm”. *IEEE Trans. Audio, Speech, Language Process*. [Online]. 38(1), 1999. [Jun. 1, 2018]

[7] Dhiman, J., et al. “Comparison between adaptive filter algorithms (LMS, NLMS and RLS)”. *International Journal of Science, Engineering and Technology Research (IJSETR).* [Online]. 2(5), 2013. [Jun. 1, 2018].