BROAD-BAND NOISE SUPPRESSION USING SHORT-TIME SPECTRAL ATTENUATION

Danilo Pesevic

MUMT 605 Final Project

Dec. 17, 2024

I. ABSTRACT

Audio recorded under nonideal conditions or with outdated equipment becomes degraded in a variety of ways. To recover this audio, one can apply digital audio restoration techniques. In this paper, an audio recovery algorithm for suppressing broad-band background noise in speech and music recordings was implemented in MATLAB. The implementation uses the short-time spectral attenuation method, based on the short-time Fourier transform and a Wiener-filter suppression rule. Testing with a range of recordings revealed that the algorithm was successful in reducing background noise, though its performance was limited by background noise amplitude and recording audio complexity.

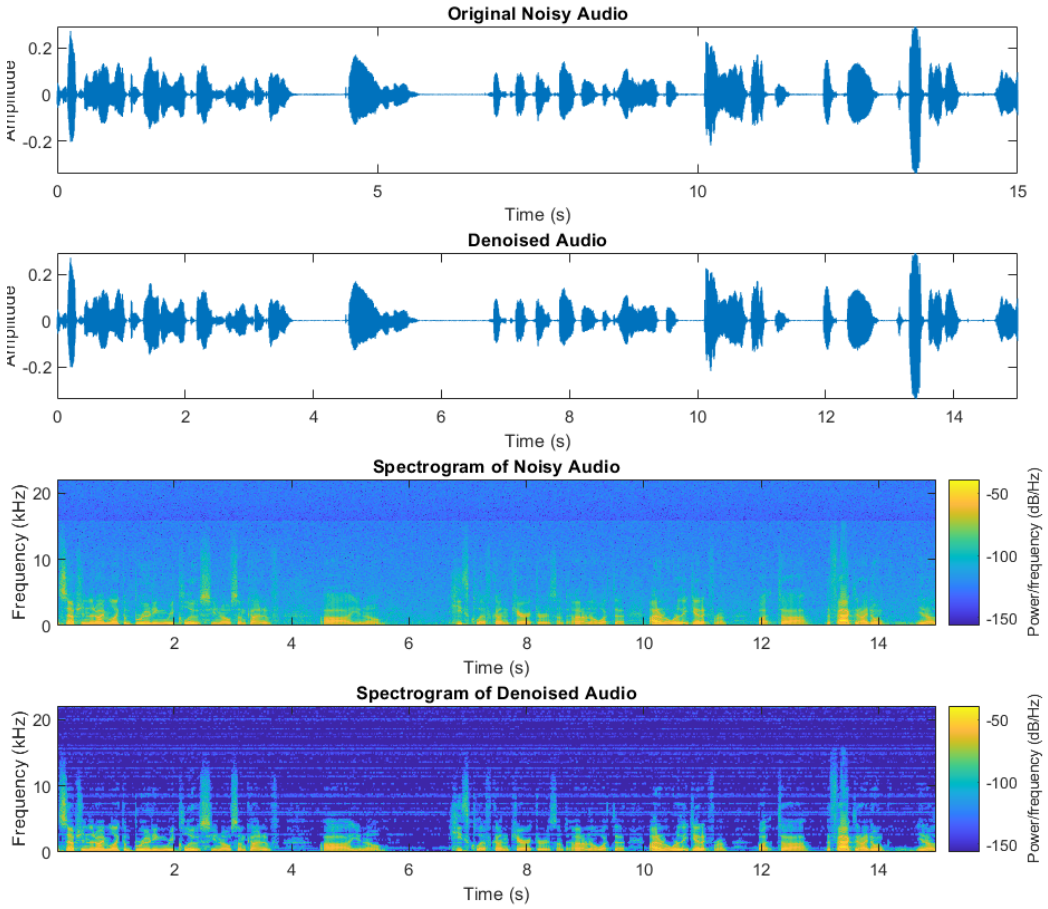

Figure 1. Time and spectrogram view of a degraded speech signal before and after denoising.

II. INTRODUCTION

Continuing advances in recording technology allow for ever-higher-fidelity audio recordings. However, there exists a wealth of audio recorded without high-fidelity equipment or under nonideal circumstances – resulting in degraded audio signals. Audio degradation comes in many forms, depending on the recording medium. These can include clicks, broad-band background noise (hiss), pitch variations (wow and flutter), and distortion, among others [1], [2]. The introduction of digital audio restoration techniques allows for "a much greater degree of flexibility" and capability compared to older, analog methods of restoration (such as splicing of magnetic tape or low-pass filtering for high-frequency background noise) [1], [2].

This paper explores removal of stationary broad-band background noise using digital audio restoration methods and implements an algorithm to suppress this noise based on the short-time Fourier transform (STFT) in MATLAB. An outline of the paper is as follows: a summary of audio degradation mechanisms and common restoration techniques will be given, followed by an explanation of the method selected for this project. The paper will finish with a discussion on the successes, limitations, and challenges of the implementation, as well as outlining some possible further work.

2.1. Motivation

Audio restoration allows us to recover and preserve historical recordings, which could include culturally significant audio materials, whether they be speech or music. In improving audio quality, restoration techniques improve accessibility through making degraded recordings more understandable. By digitizing and restoring audio recorded onto older media, such as vinyl records or cassette tapes (or even early digital recordings), we not only preserve them for archival purposes, but we also open the possibility for their use and enjoyment by new generations of listeners. By exploring a simple technique for removing background noise, a form of noise common to all forms of recording, this paper hopes to serve as an introduction to the topic, and a steppingstone for further research and contribution into audio restoration.

III. RESEARCH CONTEXT

3.1. Classes of Audio Degradation

Sources of audio degradation vary depending on recording technique and storage medium. For recordings made on wax cylinders or vinyl records, where audio is played back by passing a needle over grooves made into the material, the most common degradation is clicks – a local degradation consisting of "short bursts of interference random in time and amplitude" [2]. Their cause is imperfections, such as scratches or dirt, in the grooves of the material [1]. An example of global degradation is inconsistent motor speed during the recording of analog media, such as magnetic tapes, causing variations in the pitch of the recording. This results in a vibrato-like fluctuating sound, called "wow" at lower frequencies and "flutter" when the fluctuations are more rapid [3]. Audio distortion, another global degradation, can take many forms, such as saturation (soft clipping) in tape caused by the properties of the magnetic recording process [3],

or any other non-linear amplitude related effects [1]. This paper will focus specifically on global broad-band noise, prevalent in all forms of old recordings.

Broad-band noise, otherwise known as background noise, white noise, or hiss, can arise through a variety of mechanisms. These include electrical interference picked up by the audio transducers and “irregularities in the storage medium” [1] (such as the random unaligned magnetic domains in tape [3]). Background audio such as accidental captures of speech, or audience sounds, are not treated as “noise” here.

3.2. Audio Restoration Techniques

While this paper focuses on background noise reduction using a short-time spectral attenuation (STSA) technique, it is useful to understand other existing methods for audio restoration to better understand the context of this project.

Many existing techniques rely on statistical modeling of audio signals, such as autoregressive modelling, linear predictive modelling, sinusoidal modelling, or state-space modelling [1], [2], [4]. These models can then be used to reconstruct undegraded audio from a degraded sample. They require statistical parameter estimation, which can be done in a variety of ways. For example, in [4], a linear predictive model uses the Levinson-Durbin algorithm to estimate predictive coefficients for removing clicks. Other estimation techniques include maximum likelihood and Bayesian estimation [2].

Other techniques, including the one implemented in this paper, employ the STFT. The STFT can be used in a wide array of ways to suppress noise, though perhaps the most popular is STSA. This involves analyzing an audio signal using the STFT and attenuating each channel based on a suppression rule before resynthesizing the now denoised audio [1], [5]. For example, the authors in [6] apply Bayes prediction combined with the STFT to estimate the minimum mean-square error (MMSE) of a silent part of an audio recording, which can then be used to appropriately attenuate spectral coefficients, suppressing noise. A common suppression rule, also based on the MMSE performance criterion, is the Wiener filter (named after the mathematician and father of cybernetics, Norbert Wiener) [5]. A variation of the Wiener filter is employed for this project, drawing on the suppression rules listed in [5].

Finally, audio restoration techniques may incorporate machine learning. In [7], the authors use principal component analysis (PCA) to reduce background noise in speech recordings. After decomposing the noisy signal onto its principal components, this technique allows the audio signal to be reconstructed using only the correlated speech signals, leaving behind the uncorrelated noise signals. PCA can be combined with other techniques, such as in [8] where STSA is combined with a principle component technique, almost doubling the signal-to-noise ratio of degraded speech audio. Degraded audio may also be restored using neural networks (NNs). For example, in [9], an “audio-to-audio” NN is trained to denoise music recordings with the help of the STFT.

IV. BACKGROUND

This section serves to provide some mathematical background for the STSA audio recovery technique.

4.1. The Short-Time Fourier Transform

The STFT is a signal processing tool which provides the amplitude of a signal in terms of time and frequency [10]. Mathematically, we can define the discrete STFT X of time-domain signal x as [5]:

$$X[p,k] = \sum_{l=0}^{L-1} w[l]x[pM+L]e^{-\frac{j2\pi kl}{L}} \quad k = 0,1,2,\dots,L-1 \tag{1}$$

where p represents the frame number (related to time), k represents bin frequency, l the sample index, L the total number of samples, w the windowing function, M the hop/step size (giving an overlap of $L-M$) [5], [10].

To recover the time-domain audio signal, we can apply the inverse short-time Fourier transform (ISTFT) using the overlap-add method (OLA), where each frame is inverse transformed and summed with overlap. The inverse transformed frame x_p is given by [5]:

$$x_p[l] = g_{corr}[l] \sum_{l=0}^{L-1} X[p,k]e^{\frac{j2\pi kl}{L}} \tag{2}$$

where g_{corr} is a function correcting gain modifications during the STFT processing, given as [5]:

$$g_{corr}[l] = \frac{1}{w[l]} \tag{3}$$

so long as the overlap is less than 50%. To then recover the original signal, overlapping segments x_p are summed [5]:

$$x[n] = \sum_{p} x_p[n-pM] \tag{4}$$

for $0 \leq n - pM < L - 1$.

4.2. The Short-Time Spectral Attenuation Method

When attempting to suppress white noise in a recording, it can be treated as a stationary signal which is uncorrelated with the undegraded audio [1], [5]. Assuming the background noise is stationary/uncorrelated with the audio and additive (fair assumptions for broad-band noise), the noisy audio x can be represented as

$$x[n] = s[n] + v[n] \tag{5}$$

where s is the undegraded audio and v is the noise. By obtaining the STFT of x and of noise signal v, we can then recover s by applying a Wiener-filter derived noise suppression rule [1]. The suppression rule used in this implementation is taken from [5]:

$$G[k] = \frac{S_X[k] - S_V[k]}{S_X[k]}, \qquad S_X[k] > S_V[k] \tag{6}$$

where G is the gain applied to each frequency bin, and S represents the spectrum of a signal frame (the squared magnitude of its discrete Fourier transform). If the spectrum of the noise exceeds that of the degraded signal, then $G[k] = 0$. In effect, this suppression rule applies a Wiener filter to each frequency bin of a frame for which the signal is stronger than the noise, thereby suppressing the noise, and otherwise silences that frequency bin.

V. IMPLEMENTATION

This section summarizes the implementation of the STSA method for background noise reduction in MATLAB. A simplified flowchart diagram is given in figure 2.

An audio file must be specified at the top of the script. Stereo files are then converted to mono before being passed to the STSA processing code. The STFT and ISTFT were performed using built-in MATLAB functions `stft()` and `istft()`, respectively. `istft()` uses the OLA method as described in section III of this report. STFT frame size $L = 1024$, with overlap of 50%, and a Hann window were chosen based on trial and error, though these parameters can be easily modified.

To compute the noise spectrum, an excerpt of the degraded audio containing only noise (i.e., a silent part of the recording) must be specified. Here, these excerpts were obtained by cropping the degraded audio in Audacity and were on the order of hundreds of milliseconds in length. The spectrum of the noise is computed by averaging and then squaring its STFT magnitude across all frequency bins, giving a single frame-length vector which can be used to compute the spectral gain coefficients, G. In the script, this is accomplished in a for-loop, iterating through each frame of $X[k]$, computing its spectrum, and applying equation 6. Each G value was saved into a matrix, and thus attenuating the spectrum of the degraded audio was accomplished by multiplying its STFT

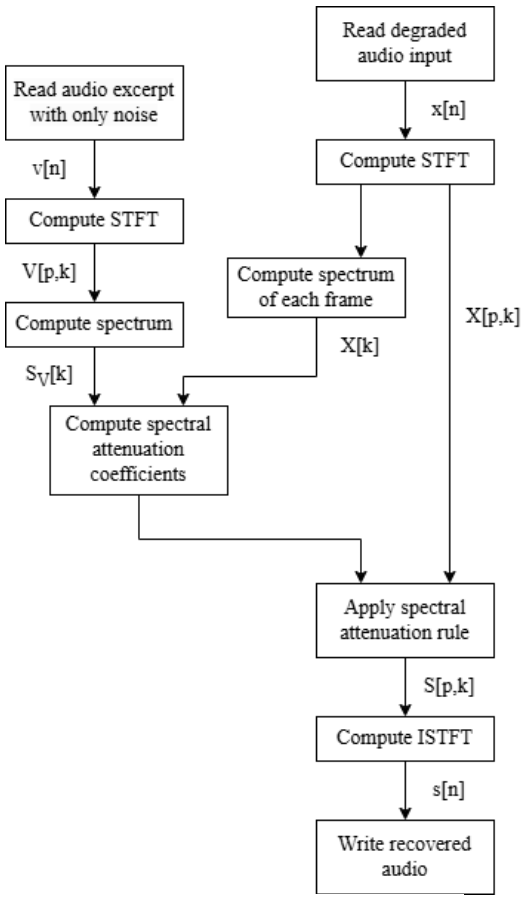

Figure 2. Flowchart of the STSA MATLAB script.

$X[p,k]$ by matrix $G[p,k]$, element by element, producing denoised spectrum $S[p,k]$. Applying the ISTFT function to $S[p,k]$ produces the denoised audio, $s[n]$.

Note that the audio signals were zero-padded before taking their respective STFTs, as otherwise the process did not work. This addition was found in the MATLAB code of [5]. Another arbitrary modification made to the algorithm was multiplying the STFT of the noise by a factor of 50, as otherwise the attenutation did not make a significant difference in noise reduction.

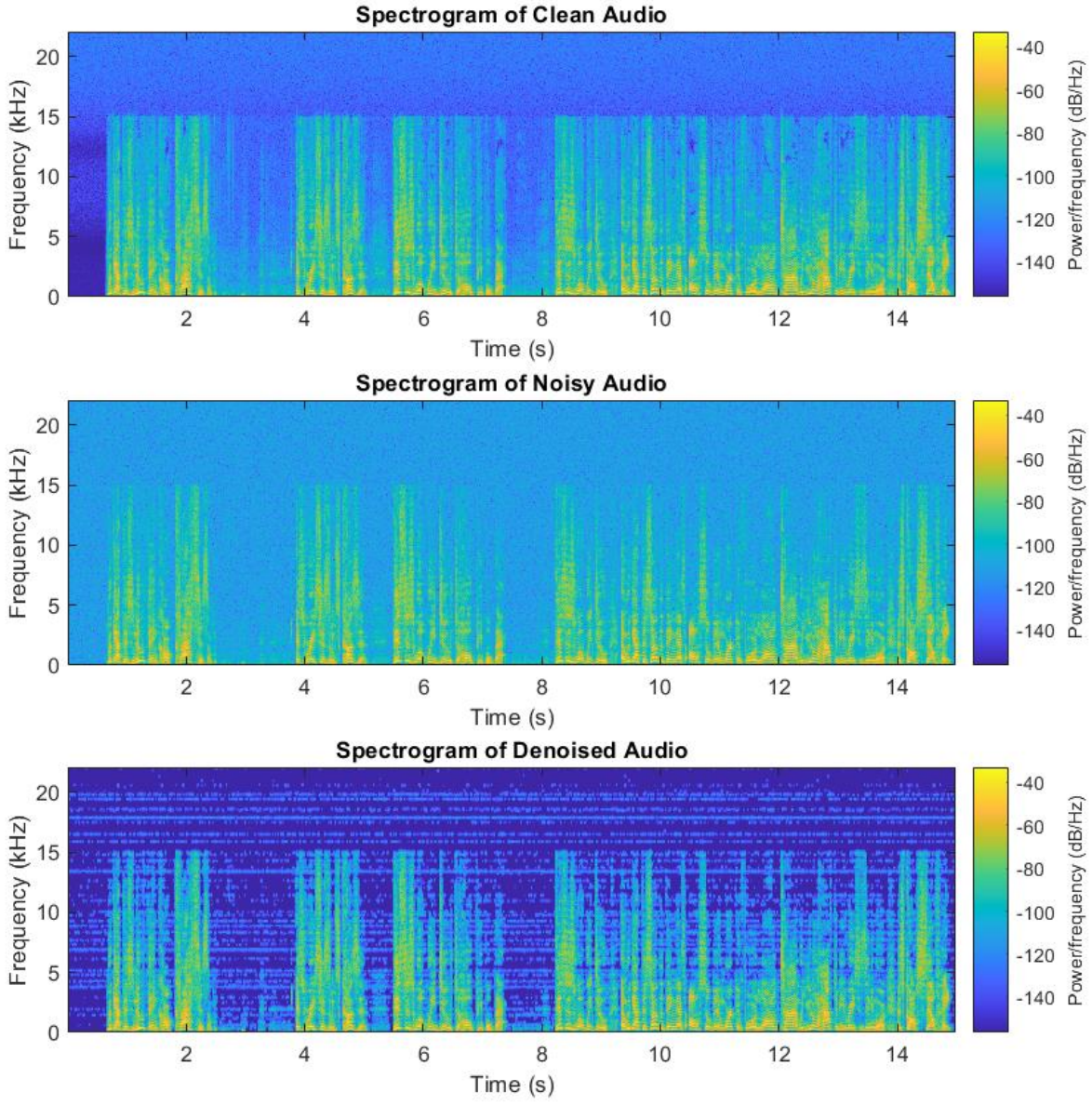

Figure 3. Spectrograms of a speech sample before and after adding noise, and then after denoising with the STSA algorithm.

VI. RESULTS

This section summarizes the results of the project and provides some discussion on the strengths and weaknesses of the project, challenges encountered during its development, and some possible future work.

6.1. Implementation Results

Ultimately, the algorithm implemented for this project successfully suppressed the background noise in degraded audio samples. Figure 3 illustrates the effectiveness of the algorithm by comparing the spectrograms of a clean speech signal, the same signal once white noise was added, and the final denoised spectrogram. It is clear in the figure that the STSA suppression method does reduce the amount of noise in the spectrogram, and this can be audibly perceived by listening to the denoised audio.

The algorithm was tested on speech and music samples, some of which already had noise while others had white noise added with a MATLAB script which combined an audio file with attenuated noise generated using the `rand()` function. As can be seen in figure 3, while the recovered audio did sound noticeably less noisy, some artifacts remained. It was observed that performance worsened with increasing noise amplitude, with more obvious artifacts present in the recovered audio.

Parameter testing seemed to show no noticeable difference in performance between the Hann and Hamming windows. Overlaps of 25%, 50%, and 75% were all tested, and it was found that 50% overlap produced the most satisfactory results. It was also found that frame sizes of 1024 and 2048 both produced acceptable results, though the latter left more unpleasant artifacts when used on noisier audio. Other frame sizes did not produce acceptable sounding denoised audio.

By subtracting the recovered audio from the original sample, the residual noise signal was examined. It was found that even in the best sounding examples, the residual still contained more than only noise – it contained “musical noise,” which is expected for the STSA method [1]. This musical noise was more noticeable in the residuals of music samples than in speech samples and became more prominent as the noise volume increased.

The algorithm seemed to work better for speech recordings than for music, though the music results were mostly satisfactory. There was a greater tolerance for noise volume when processing speech, with speech samples more successfully denoised (less artifacts) when compared to music samples with comparably loud noise.

6.2. Discussion

The STSA algorithm implemented in this paper shows the potential of digital audio restoration techniques for recovering degraded recordings. Despite the relative simplicity of STSA, the results show a noticeable reduction in broad-band noise for both speech and music samples, as evidenced by the spectrograms in figures 1 and 3. More broadly, this project highlights the power of the STFT as a tool in audio signal processing tool.

A significant challenge I faced in the implementation of the STSA algorithm was translating the algorithm from its description in [1] and [5] to a MATLAB script. Neither work gives a particular detailed description of an actual implementation, and thus I spent a considerable amount of time trying to develop the code. For example, I did not realize that that every frame was to have its own vector of spectral gain coefficients, as opposed to a single G vector for each frame. I resolved this challenge by using [10] to understand the mathematics behind the algorithm at a lower level, and this allowed me to successfully implement the STSA algorithm. Another challenge was establishing the ideal algorithm parameters, which was accomplished with a subjective trial and error testing procedure. It is possible that there are more optimal combinations of parameters that I did not test. Even once acceptable STFT parameters were selected and the algorithm was correctly implemented, the results were not satisfactory until the noise STFT amplitude was multiplied by a factor of 50. This seemed arbitrary, though I could not find any alternative or any explanation in the literature. The requirement for zero-padding input audios before taking their STFTs was also not mentioned in the sources explaining the STSA method, yet the algorithm did not function without this being done. This was a challenge only resolved when I found that zero-padding was necessary in the appendix of [5].

While the STSA technique was generally effective, it did not come without limitations. The presence of artifacts, particularly the musical noise present in the residual noise audio, illustrates one of the major drawbacks of STSA-based denoising. Higher noise levels produced more artifacts, which suggests that this method is not effective for the recovery of very degraded audio. Additionally, the reliance on a priori noise estimates is a disadvantage of this implementation. Having to isolate a silent portion of the degraded audio is not only tedious but may be impossible for some recordings. This implementation was also limited to only removing global, broad-band noise, and does not suppress the other forms of audio degradation such as clicks or wow/flutter. Additionally, at this stage, the algorithm only works in mono.

Despite its limitations, the results of this project are quite satisfactory and highlight the effectiveness of STFT-based background noise suppression, especially for degraded audio with lower noise amplitude. This project also helped me to better understand and familiarize myself with the STFT as a signal processing tool and served as an introduction to digital audio restoration.

6.3. Future Work

While functional in its current state, this project has much room for improvement before becoming a truly useful audio restoration tool.

The inherent limitations of the STSA method (e.g., musical noise) should be remedied through alternative suppression rules, such as the Ephraim-Malah rule [1], [5], [6], perceptual noise reduction criteria [1], or through combining the STSA method with machine learning techniques like PCA [8]. Alternatively, the STSA method could be replaced entirely with a machine learning based technique such as using a neural network [9]. It appears that these sorts of techniques are becoming more dominant in audio restoration.

To improve the project in its current state would require a more in-depth and systematic analysis of the STFT parameters and how they affect the results. Testing on a greater number of audio samples would also help fine-tune the algorithm, improving the results. Developing an adaptive noise estimation technique, removing the need for the a priori noise sample, would improve the usability of this project and could help expand it to real-time use.

Additionally, the project could be expanded to address other forms of audio degradation.

VII. CONCLUSION

This paper explored digital audio recovery by implementing a STSA algorithm for broad-band noise suppression in degraded audio signals, using the STFT. The project successfully reduced background noise in both music and speech samples, however limitations such as the presence of musical noise artifacts and a dependency on a priori noise samples underscore areas for future work.

Through testing on speech and music samples of varying noise levels, it was found that the implementation is best suited for lower noise amplitude speech signals, because of increased artifacts in denoised complex musical or noisier signals.

Further developments could include using a different noise suppression rule, supplementing the STSA method with another technique such as PCA, or replacing the STSA method entirely in favor of a machine learning model. Additionally, this project would benefit from more systematic testing of its parameters, and from the development of automatic noise spectrum estimation.

Despite its limitations, this project highlights the power of STFT-based audio recovery and serves as an introduction to digital audio restoration techniques more broadly.

REFERENCES

[1] S. Godsill, P. Rayner, and O. Cappé, 'Digital Audio Restoration', in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds. Boston, MA: Springer US, 2002, pp. 133–194.

[2] S. J. Godsill and P. J. W. Rayner, 'Introduction', in *Digital Audio Restoration: A Statistical Model Based Approach*, London: Springer London, 1998, pp. 1–11.

[3] M. Camras, *Magnetic Recording Handbook*. New York, NY, USA: Van Nostrand Reinhold Company, 1988.

[4] D. Kawano, T. Ogawa, and H. Matsumoto, 'A proposal of the method to suppress a click noise only from an observed audio signal', in *2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, 2017, pp. 93–96.

[5] V. Hella, "Digital audio restoration: denoising photograph recordings," master's thesis, Norwegian University of Science and Technology, Trondheim, Norway, 2013.

[6] Z. Brajević and A. Petošić, 'Signal denoising using STFT with Bayes prediction and Ephraim-Malah estimation', in *Proceedings ELMAR-2012*, 2012, pp. 183–186.

[7] B. Li, 'A Principal Component Analysis Approach to Noise Removal for Speech Denoising', in *2018 International Conference on Virtual Reality and Intelligent Systems (ICVRIS)*, 2018, pp. 429–432.

[8] O. Julius, I. C. Obagbuwa, A. A. Adebiyi, and E. B. Michael, 'Implementation of Audio Signals Denoising for Perfect Speech-to-Speech Translation Using Principal Component Analysis', in *2023 International Conference on Science, Engineering and Business for Sustainable Development Goals (SEB-SDG)*, 2023, vol. 1, pp. 1–6.

[9] Y. Li, B. Gfeller, M. Tagliasacchi, and D. Roblek, 'Learning to Denoise Historical Music', *arXiv [eess.AS]*. 2022.

[10] J. O. Smith, *Spectral Audio Signal Processing*, 2011 ed. [E-book]. Available: http://ccrma.stanford.edu/~jos/sasp/.