CLICK REMOVAL IN DEGRADED AUDIO

Report for Module EEP55C22 Computational Methods

Dian Zhuang, Student ID 22329261
Trinity College Dublin
zhuangd@tcd.ie

October 2022

This report is submitted in part fulfilment for the assessment required in EEP55C22 Computational Methods. I have read and I understand the plagiarism provisions in the General Regulations of the University Calendar for the current year. These are found in Parts II and III at http://www.tcd.ie/calendar.

This report describes the algorithm designed for detection and removal of clicks in archived audio tracks. The sound manifests as a short sharp click or thump in the audio track. The generation of these degradations may be viewed from two perspectives: in commercial recording, the degradation could be an artifact of the recording (all kinds of analog recording display hiss noise), or it may be the consequence of media corruption (eg. scratches on a phonograph). Degradation of the audio signal during personal recording may be caused by restrictions of the generally accessible recording medium, limitations of inexpensive recording equipment, or mistakes in recording device operation.[2] For Our problem is to find an effective technique for the restoration of audio signals degraded by impulsive clicks or by broadband hiss(More attention will be paid here to clicks).

1 Autoregressive (AR) modeling

The autoregressive (AR) model is one of the fundamental models we shall employ. An audio signal is modeled as a linear white noise process applied to a signal. A white noise process is subjected to a time-invariant all-pole filter. Each output sample xn of the output process can be considered to be the weighted sum of a limited number of previous samples plus a single sample en from the random white noise process.[2] The order of the AR process x is denoted by P. The AR coefficients are defined as the P coefficients ai. The all-pole filter's transfer function is:

$$X_(n) = \sum_{i=1}^{p} a_{(i)} x_{(n-1)} + e_{(n)} \tag{1}$$

$$H_{(z)} = \frac{1}{A_(z)} \tag{2}$$

where

$$A_{(z)} = 1 - \sum_{i=1}^{p} a_{(i)} z^{(-i)} \tag{3}$$

The white noise process is known as the invention or stimulation process because of its involvement in the intellectual synthesis of the AR process. When equation (1) is thought to be a predictor, the prediction error sequence e_i is commonly utilised.[2]

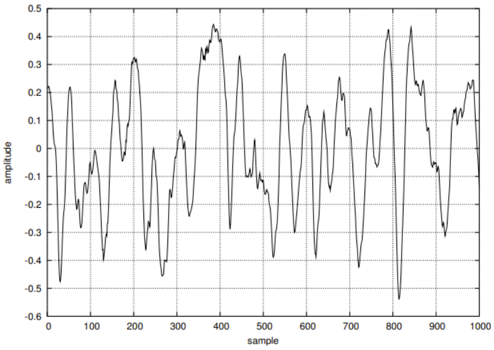

Figure 1: *Example audio waveform [2]*

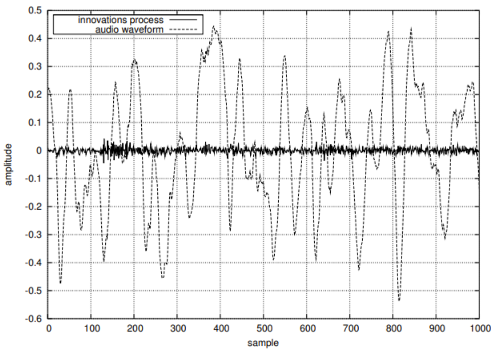

Figure 2: *Audio waveform and innovation sequence[2]*

A rough estimate of a can be derived by minimising e. Meaningful e reduction. This is referred to as covariance estimation.

$$a = (G^T G)^{-1} G^T X_1 \tag{4}$$

2 Removal of Clicks

Clicks are localized bursts of impulsive noise that occur in audio transmission. This section delves into time series modeling of audio signals, click noise detection, and click noise reduction. Click removal for both simulated and actual corruption is proven. We utilise formulae and symbols throughout this section.

2.1 Modelling of clicks

We will focus on the additive model here, while many of the conclusions are at least resilient to replacement noise. An additive model for localized deterioration is as follows:

$$y_t = x_t + i_t n_t \tag{5}$$

When x_t is the underlying audio signal, n_t is a corrupting noise process, and it is a 0/1'switching' process that takes the value 1 only when the localized deterioration is present.[1]

2.2 Interpolation of missing samples

Most click-removal approaches rely on interpolation to replace missing or damaged data with estimations of their real value. It is typically safe to assume that clicks haven't messed with the timing of the content, thus the objective is to fill in the 'gap' with relevant material of the same time as the click. We begin by presenting a generic framework for the interpolation of Gaussian signals.

2.3 Autoregressive model-based interpolation

For the general Gaussian case, we represent the algorithm in matrix/vector notation, where the position of the missing sample can be arbitrarily specified in the data block by detecting the switching vector i. The method can be derived from the least squares (LS) method and the location of the missing sample is specified by the detection switch vector i. The method can be derived from the least squares (LS) or maximum a $posteriori$ (MAP) principles. Here the focus is more on least squares (LS). Consider a set of data samples x obtained from an AR process with parameters a. In terms of the data vector, we may express the excitation vector e.

$$e = Ax = A(Ux_{(i)} + Kx_{-(i)}) \tag{6}$$

now define $A_{(i)} = AU$ and $A_{-(i)} = AK$, to give

$$e = A_{-(i)}x_{-(i)} + A_(i)x_(i) \tag{7}$$

The sum squared of the prediction errors over the whole data block is

$$E = \sum_{n=P+1}^{N} e_n^2 = e^T e \tag{8}$$

E can be seen as a measure of how well the data 'fit' the AR model and can be extended and differentiated using the standard vector-matrix calculus to find its minimum value.[1]

$$\frac{\partial E}{\partial x_(i)} = 2e^T \frac{\partial e}{\partial x_(i)} = 2(A_{-(i)}x_{-(i)} + A_{(i)}x_{(i)})^T A_{(i)} = 0 \tag{9}$$

Hence, we obtain

$$X_(i) = -(A_{(i)}^T A_{(i)})^{(-1)} A_(i)^T A_{-(i)}x_{-(i)} \tag{10}$$

2.4 Detection

We developed models that fit clean audio signals. The following step is to locate corrupted samples inside noise-degraded audio.

2.4.1 Parameter estimation

To improve detection localisation, the forward and backward prediction error sequences can be combined. The smallest amplitude prediction error between forward and backward values at a given sample is one basic way. The notion is that corrupted samples should have a large amplitude in both sequences, but smearing should present in just one of them at a given place.[2]

2.4.2 Noise identification

In an actual inspection system, it is vital to confirm that each sample has been corrupted. To do this, we employ a threshold to identify samples when the prediction error reaches a given amount. The threshold is set in relation to the standard deviation of the sequence of prediction errors in a block. The standard deviation of a block's prediction error sequence (or a robust estimate of the uncorrupted standard deviation).

3 Matlab implementation flow

1. Audio read

2. Data block
3. Interative click detect
4. Calculate residuals
5. Calculate AR parameters from residuals
6. Estimate click locations
7. Interpolate unknown volues
8. Estimate AR parameters
9. Replace corrupted data with clean interpolated data

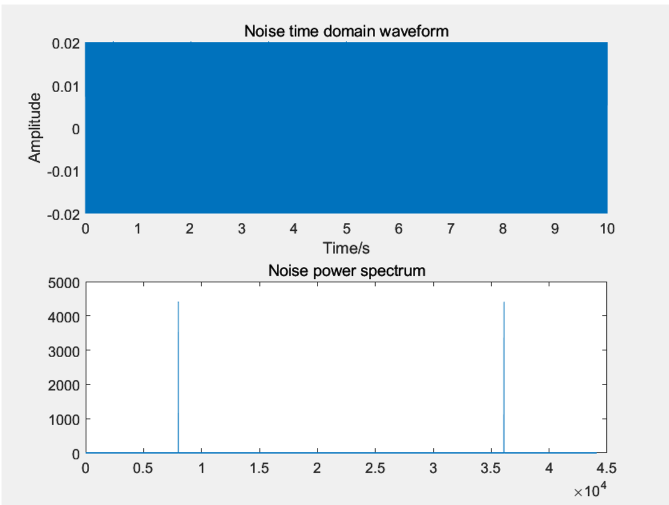

Figure 3: *figure contains two images the upper image is the Noise time domain waveform the lower image is the Noise power spectrum*

A Future Directions

In this assignment we are using a recorded piece of audio or adding noise or clicks to a clean piece of audio after editing, so we can take a longer view and see that the future of audio noise removal must be in real-time audio and video, such as real-time audio echo cancellation and noise cancellation, which involves the knowledge of physical acoustics is of interest to the industry.

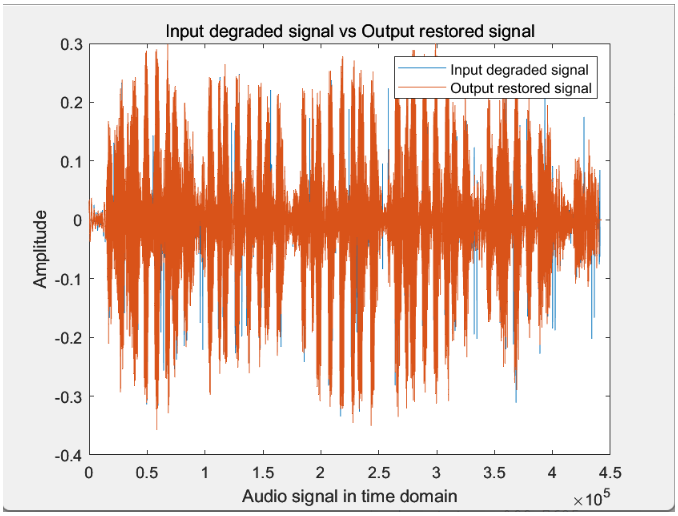

Figure 4: *Input degraded signal vs Output restored signal*

Bibliography

[1] GODSILL, S. H., AND RAYNER, P. J. W. Digital audio restoration: A statistical model based approach.

[2] NUZMAN, J. Audio restoration : An investigation of digital methods for click removal and hiss reduction.

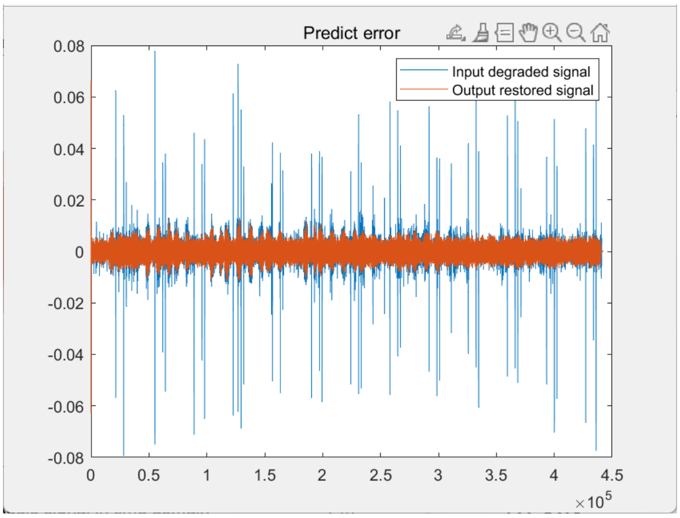

Figure 5: *Prediction error of the input attenuated signal compared to the output recovered signal*