Active Noise Control of Speech in Headphones using Linear Prediction

Mikkel Krogh Simonsen, Kasper Kiis Jensen, Maxime Démurger,
Oliver Palmhøj Jokumsen, Christian Claumarch
Aalborg University, Acoustics and Audio Technology - 1st Semester

Active Noise Control (ANC) is a widely used technique. However most solutions are targeted at attenuating static noise with limited bandwidth. This paper focusses on using prediction to improve performance for non stationary signals. The solution proposed uses an FIR-filter adapted by an Filtered-x Least Mean Square (FXLMS)-algorithm combined with Linear Prediction (LP) to attenuate a speech signal. This solution uses the quasi-periodic properties of speech to estimate the upcoming content of a speech signal. Thereby decreasing the sample delay in the system which increases bandwidth, yielding better attenuation.

1 Introduction

Active Noise Control (ANC) is a used method in a wide variety of applications. Research is being done in creating silents zones[1], improving room acoustics using multiple loudspeakers[2] and is already available in various headphones.

Noise canceling headphones are used in various environments which have different noise characteristics. These characteristics vary from periodic low frequency noise (10-200 Hz), e.g. from machinery and helicopter rotor[3], mid-range frequency noise (200-4000 Hz), e.g. speech[4], to high frequency noise (4-20 kHz), e.g. turbine noise[3]. The high frequency noise is naturally attenuated by using a closed headphone cup[5] and static low frequency noise are attenuated by the present consumer headphones[5].Speech can be shown in figure 1, not to be attenuated by active noise cancellation in present market ANC headphones.

Here is going to be a graph of different headphones tested with non stationary signals (Speech)

Figure 1: It is clearly seen that speech is not attenuated by current headphones.

This paper will be focusing on attenuation of speech using an ANC-system. Speech will hence be seen as the primary noise source. The solution will be based on a digital feedforward system using the Filtered-X Least Mean Squares (FXLMS). The solution is chosen due to it being the optimal system for non stationary signals[6]. The general solutions in ANC are described thoroughly by Hansen et al. [7].

The problem of a feedforward system, shown in figure 2, is the dependency of noise having a larger propagation delay from the noise is measured (1) to a headphone loudspeaker(2) can produce a counter-phase signal in the desired point(3).

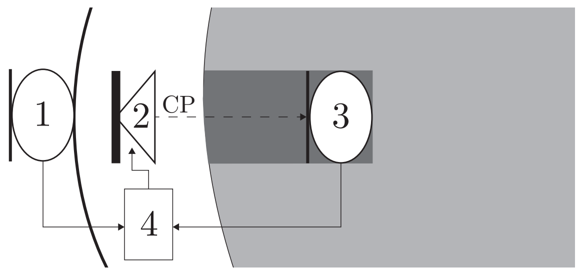

Figure 2: Simplified circumaural ANC Headphone fitted on ear. Showing the reference microphone (1), a headphone loudspeaker (2), an error microphone (3) and a DSP (4).

When implemented in a real-time system delay will occur, introduced by sampling and processing of the signal. For instance, measured on a cost-wise ΣΔ-converter a sampling/conversion delay in the range of 220 to 880 μs is introduced, depending on a sampling frequency of 192 kHz to 48 kHz respectively. Assuming a 0° angle of incident, which is the ideal case, the reference microphone must be placed 75.5 mm to 302 mm further out from the error microphone, to compensate for the delay. In order to compensate for the introduced delay a Linear Prediction (LP)-scheme is proposed as a solution. In order to do prediction, certain signal characteristics must be known. Prediction of speech is described by Wai C. Chu [8].

The paper is split into two parts. The first part describes the feedforward FXLMS system and how to predict speech using linear prediction. The second part describes simulations and the implementation. Performance of the feedforward FXLMS system with and without prediction are determined and compared.

2 Methods

Using Figure 2 as the outline, the DSP(4) is expanded into Figure 3.

2.1 Feedforward ANC using FXLMS

The system in Figure 3 outputs a control signal $y[n]$, which ideally is a counter-phase signal of the noise. The signal is generated by inputting the reference signal $x[n]$ into a control filter consisting of adaptive coefficients $b[z]$ representing the transfer function from the reference microphone to the headphone loudspeaker. The coefficients are adapted using the FXLMS algorithm. The FXLMS algorithm inputs the filtered reference $f[n]$ signal along with the error signal $e[n]$. The filtered reference signal is used combined with the error signal to determine new optimal coefficients for the control filter, this is shown in equation 2. The signals from (1)(3) are converted to the digital domain using an ADC and anti aliasing filters (AA) before processing. When processed, the output $y[n]$ is reconstructed and converted using a DAC.

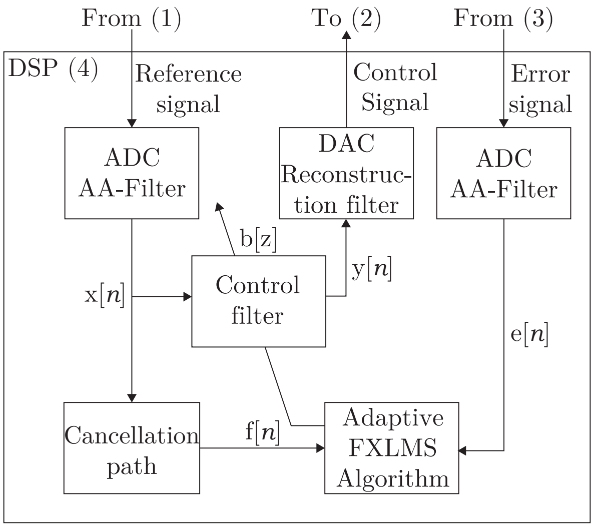

Figure 3: Adaptive feedforward ANC system.

Control Filter The filter, shown in equation 1, is initialized with the inverse of the measured impulse response of the represented transfer function. An order of 256 taps is chosen based on subjective testing in simulation.

$$y[n] = \sum_{n=0}^{L-1} b_j x[n-j] \tag{1}$$

Where:

$b_j[n]$ is the weight coefficients written as $b[n] = [b_0[n], b_1[n], \cdots, b_{L-1}[n]]^T$

FXLMS is the optimization algorithm which updates the control filter coefficients using the FXLMS method shown in Equation 2.

$$b_j[n+1] = b_j[n] - 2\mu e[n] f[n-j] \tag{2}$$

Where:
μ is the convergence factor

$e[n]$ is the error

$f[n]$ is the reference convolved with the Cancellation Path

Cancellation Path (CP) is the transfer-function from the headphone loudspeaker to the error microphone. In the literature [7] the CP is adaptively adjusted, but it is assumed constant because the position of the headphone does not change while measured on a Head and Torso Simulator (HATS). This assumption is made because it is irrelevant for verifying if LP is a plausible solution.

2.2 Characteristics of Speech

Speech can be split into two main classes, voiced and unvoiced. Voiced sounds are characterized by a strong periodicity, with the fundamental frequency referred to as the pitch frequency (50 - 500 Hz). Unvoiced sounds are characterized as random. Speech is a non stationary signal and can only be assumed Wide Sense Stationary (WSS) for periods of 20 - 30 ms[8].

2.3 Linear Prediction of Speech

In order to predict future samples the Auto Correlation Function (ACF), used in LP, of speech must be estimated in frames. The outlines of the prediction system is shown in figure 4.

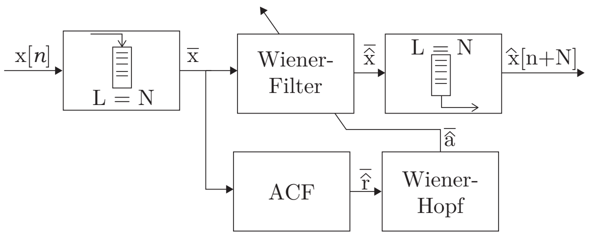

Figure 4: Linear prediction system.

Utilizing a nonrecursive estimation of the ACF, shown in Equation 3, it is possible to determine the linear prediction coefficients $\bar{\hat{a}}$ (LPCs) used for predicting future samples [9]. When estimating the ACF of the input (x) a Hamming window (w) is applied.

$$\hat{r_x}[l] = \sum_{n=|l|}^{N} x_l[n] w_l[N-n] \tag{3}$$

Where:
$x_l[n] \; = x[n]x[n-l]$

$w_l[n] \; = w[n]w[n+l]$

l is the lag

N is the frame size

The LPCs are determined using Equation 4, known as the Wiener-Hopf equation.

$$\hat{R}\bar{a} = -\bar{r_x} \tag{4}$$

Where:
$\hat{R}$ is the covariance matrix $\hat{C_{xx}}$

$\bar{\hat{a}}$ is the LPCs, $\bar{\hat{a}} = [\hat{a_1}, \hat{a_2}, \cdots, \hat{a_N}]^T$

$\bar{\hat{r_x}}$ is the ACF, $\bar{\hat{r_x}} = [\hat{r_x}[1], \hat{r_x}[2], \cdots, \hat{r_x}[N]]^T$

Equation 4 can be rewritten as shown in Equation 5 yielding the LPCs directly.

$$\bar{\hat{a}} = -\hat{R}^{-1}\bar{\hat{r_x}} \tag{5}$$

Calculating $\hat{R^{-1}}$ is computationally heavy on a DSP. To estimate the LPCs the Levinson-Durbin method is used [9]. Prediction using Wiener filtering, shown in equation 6, can then be applied to the current frame for prediction of the next frame.

$$\hat{x}[n+1] = -\sum_{i=1}^{N} \hat{a}_i x[n-i] \tag{6}$$

Using equation 6 in cascade where $\hat{x}[n+2]$ is estimated using $\hat{x}[n+1]$ and $x[n]$. The predicted frame is then used as input for the ANC system.

Results

At this point in time all results are not yet certain, however we will give an idea of how they are going to turn out. As presented in the paper thus far, we know what kind of noise we would like to cancel out and how we want to do it. The following list tells which graph will be used to show the results of the acceptance tests.

- Prediction gain - The difference between the input signal and the estimated signal, measured in dB. Used in the LP part.

- A plot comparison of frequency response between: A pure signal, a signal with FXLMS noise attenuation and a signal with LP combined with FXLMS attenuation

- An expansion of Figure 5: As of now only attenuation is showed, but in the future another graph in the figure will show the attenuation of the FXLMS combined with LP - which should yield better attenuation at larger delays.

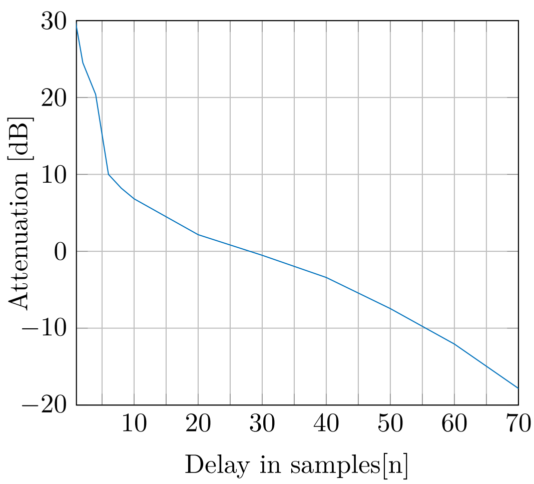

Figure 5: Attenuation achieved by the system for different system delays.

3 Dicussion

4 Conclusion

Bibliography

[1] I. T. A. . W. H. Abdulla, "Adaptive signal processing algorithms for creating spatial zones of quiet," 2014.

[2] S. B. N. . A. Celestinos, "Active low frequency sound field control in a listening room using controlled acoustics bass system will also reduce the sound transmitted to neighbour rooms," 2012.

[3] S. J. Thomas L. Lago and P.-A. Hellstrom, "Analysis of helicopter sound for the development of a new generation active headset," 1997.

[4] S. Vajrala Mangamma, "Noise cancellation of speech signal by using adaptive filtering with averaging algorithm," 2014.

[5] A. C. G. . D. M. M. Tom Frank, "Hearing thresholds, threshold repeatability, and attenuation values for passive noise-reducing earphone enclosures," 1997.

[6] C. Hansen, *Understanding Active Noise Cancellation.* Spon Press, 1st ed., 2003.

[7] D. M. . S. S. . X. Q. . L. B. . C. Hansen, *Active Control of Noise and Vibration.* CRC Press, 2nd ed., 2012.

[8] W. C. Chu, *Speech Coding Alorithms.* Wiley, 1st ed., 2013.

[9] K. M. I. K. M. N. K. F. K. L. U. S. S. A. G. D.-W. Khan, N., "Implementation of linear prediction techniques in state estimation," 2013.