

# Restoration of Howling Corrupted Audio Signals through Frequency-Domain Autoregressive Interpolation

Daniele Giacobello

December 8, 2008

## 1 Introduction

In any audio system involving simultaneous sound recording and reproduction, the coupling between the loudspeakers and the microphones can lead to instabilities which result in annoying howling sound. This problem is also known as feedback problem. It is even more critical in hearing aid where the distance between the receiver (loudspeaker) and the microphone is small. Hearing aids with open fitting have been used more commonly during the past years as they can improve comfort. They are also more sensitive to feedback problem which is limiting the possible hearing aid amplification gain. There is two different approaches to solve this problem. The first one is to prevent feedback effects, using techniques based on acoustic path identification [1]. Another solution is to cancel the howling when it is occurring. The system proposed here is based on the second approach.

The first step is to detect if howling is present. When howling appears, the system is oscillating. Works have been done to track oscillation mainly using zero-crossing rate [2, 3] or adaptive notch filtering [4]. The method presented here is based on power peak to average ratio in frequency bands. To make the detection more robust, it is used jointly with an hangover scheme.

Once the howling is detected it has to be attenuated, one straight forward method is to use notch filters. This processing is degrading the signal in the band where the filter is canceling the howling component. Therefore the filtering is combined with audio restoration process.

This report present the steps of the howling attenuation and successive audio restoration.

## 2 Attenuation

In order to guarantee an efficient restoration of the corrupted signal, a first attenuation of the howling effect is necessary. This is done since in a latter stage we are processing the audio signal employing the analysis-modification-synthesis (AMS) framework using discrete short-time Fourier transform (DSTFT). The DSTFT is particularly sensitive to the howling problem. In particular, due to the small windows used (in order to keep the algorithmic delay low) that have low spectral resolution, the howling frequency tend to mask a large portion of the spectrum making the restoration process hardly feasible. The attenuation of the howling takes effect using a notch filter. The general form of a second-order notch filter is:

$$G(z) = \frac{1 - 2 \cos(\omega_0)z^{-1} + z^{-2}}{1 - 2p \cos(\omega_0)z^{-1} + p^2 z^{-2}}, \quad p < 1, \quad (1)$$

where  $\omega_0$  is the frequency that we wish to reject and  $p$  is a control parameter that influence the band  $B$  around  $\omega_0$  where the notch filter operates. The following approximation holds well:

$$B \approx 2 - 2p \text{ [rad/s]}, \quad (2)$$

and it is easy to see that for  $p \rightarrow 1$  we have  $B \rightarrow 0$ . The howling effect has important properties that we wish to exploit in order to attenuate it efficiently. It was noted from initial preliminary analysis of the howling phenomenon that the frequency estimator was highly attracted to the low frequency region especially in speech signals. In order to avoid this problem, we restrict the estimation of the howling frequency to frequencies above 500 Hz [5]. It is also important to notice that the when the howling effect is taking place, the howling frequency does not drift, therefore not requiring a constant adaptation of the taps. Considering this, the general structure of the adaptive notch filter [6] has been modified to fit our needs. Also the fuzzy detection presented in the previous section gives us a likelihood parameter  $\alpha$  that we also wish to exploit. We divide our scheme into two parts a howling frequency detection and a filtering part.

## 2.1 Frequency Estimation

The frequency estimation is done using the method in [7] on a  $N$  samples audio segment in order to have at the instant  $n$  a frequency estimation based on the current sample  $x(n)$  and the vector of  $N$  previous samples  $[x(n - N - 1), \dots, x(n - 1)]$  requiring an initial buffer of  $N - 1$  samples. The signal is masked using the  $N$  fuzzy decision  $[\alpha(n - N - 1), \dots, \alpha(n)]$  in order to put more emphasis on the corrupted parts.

## 2.2 Notch Filtering

Once we have an estimation of the howling frequency, we can just plug the parameters  $\omega_0(n)$  and  $p(n)$  into the equation (1):

$$y(n) = x(n) - 2 \cos(\omega_0(n))x(n - 1) + x(n - 2) + 2p(n) \cos(\omega_0(n))y(n - 1) - p(n)^2y(n - 2). \quad (3)$$

Here too we can use the fuzzy detector in order to improve the overall efficiency of the algorithm, since the estimation in the initial part of the howling may not be totally accurate. The parameter  $p(n)$  is related to the band of the notch filter, therefore the stronger the howling, the wider the band has to be. Bounding the value of  $p$  between 0.6 and 0.95 and knowing that  $\alpha$  is always between 0 and 1, we can find a function that relates the two parameters:

$$p(n) = f(\alpha(n)). \quad (4)$$

A linear function has been used for this purpose. Furthermore, we can modify the output of the filtering by weighting the output of the notch filter in respect to its howling likelihood:

$$u(n) = (1 - \alpha(n))x(n) + \alpha(n)y(n). \quad (5)$$

It is clear that when  $\alpha(n) = 0$  no attenuation takes place. Also, when  $\alpha = 1$  for more than 200 samples (10 ms at 20 KHz) we interrupt the adaptation considering this time sufficient in order to find an accurate estimation of  $\omega_0$ .

## 3 Audio Restoration

The audio restoration part is based on autoregressive (AR) model-based interpolation. This interpolation procedure has proved highly successful in the

time domain restoration problems [8, 9]. In our problem we apply it in the frequency domain by interpolating the missing part of the spectrum that has been corrupted by howling. Since we have already applied a notch filter in order to reduce the howling effect, we can easily identify the part that has to be substituted as we know the central frequency  $\omega_0$  and the band  $B$  of the notch filter. The steps of the restoration process are the following:

1. Consider a segment of  $N$  samples of windowed audio data coming out of the notch filter, forming a vector  $\mathbf{x} = [x(n), \dots, x(n + N - 1)]$ , consider  $\mathbf{X}$  its Discrete Fourier Transform over  $N_{DFT} = N$  samples and the partition of half of the DFT coefficients due to the conjugate symmetry  $\mathbf{X}_p = [X(1), \dots, X(N/2)]$ .
2. Consider the notch filter parameters: the central frequency of  $\omega_0$  and its bandwidth  $B = 2 - 2p$  and transform them in the DFT domain:

$$\begin{aligned}\omega_d &= \lceil \frac{\omega_0}{\pi} \cdot \frac{N}{2} \rceil; \\ B_d &= \lceil \frac{B}{\pi} \cdot \frac{N}{2} \rceil.\end{aligned}\tag{6}$$

Knowing these two parameters, we can find the starting and ending indexes of the howling degraded part of the DFT spectrum  $i_b = \omega_d - B_d/2$  and  $i_e = \omega_d + B_d/2$ .

3.  $\mathbf{X}_p$  is then partitioned into three sections: the corrupted section that we will consider as unknown  $\mathbf{X}_u = [X_p(i_b), \dots, X_p(i_e)]^T$ , the samples to the left of the gap  $\mathbf{X}_{kl} = [X_p(1), \dots, X_p(i_b - 1)]^T$  and the remaining known samples to the right of the gap  $\mathbf{X}_{kr} = [X_p(i_e + 1), \dots, X_p(N/2)]^T$ :

$$\mathbf{X}_p = [\mathbf{X}_{kl}^T \quad \mathbf{X}_u^T \quad \mathbf{X}_{kr}^T]^T.\tag{7}$$

We then form a single vector of known samples  $\mathbf{X}_k = [\mathbf{X}_{kl}^T \quad \mathbf{X}_{kr}^T]^T$  and express  $\mathbf{X}_a$  in terms of its known and unknown components as:

$$\mathbf{X}_p = \mathbf{K} \mathbf{X}_k + \mathbf{U} \mathbf{X}_u,\tag{8}$$

where  $\mathbf{U}$  and  $\mathbf{K}$  are the “rearrangements” matrices.

4. Consider now the data samples  $\mathbf{X}_p$  as generated by an AR process with parameters  $\mathbf{a}$ , we can rewrite the excitation vector as:

$$\mathbf{e} = \mathbf{A} \mathbf{X}_p = \mathbf{A}(\mathbf{K} \mathbf{X}_k + \mathbf{U} \mathbf{X}_u),\tag{9}$$

where:

$$\mathbf{A} = \begin{bmatrix} -a_P^* & \cdots & -a_1^* & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -a_P^* & \cdots & -a_1^* & 1 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \cdots & 0 & 0 & -a_P^* & \cdots & -a_1^* & 1 & 0 & 0 \\ 0 & \cdots & 0 & 0 & -a_P^* & \cdots & -a_1^* & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & -a_P^* & \cdots & -a_1^* & 1 \end{bmatrix}.$$

Note that  $\mathbf{e}$  is not the interpolation error, but an estimate of the underlying excitation signal. It is therefore important to define properly the statistical properties of this excitation, i.e. the error criterion in the minimization process. It has been shown that the complex Laplacian probability density function (pdf) is more desirable in order to describe the DFT coefficients rather than a Gaussian pdf (see, for example, [10]). According to this, the maximum *a posteriori* (MAP) approach is done on the following distribution:

$$p(\mathbf{X}_u | \mathbf{X}_k, \mathbf{a}) \propto \exp(-\|\mathbf{e}\|_1) = \exp(-\|\mathbf{A}(\mathbf{K} \mathbf{X}_k + \mathbf{U} \mathbf{X}_u)\|_1) \quad (10)$$

$$\mathbf{e} \in \mathbb{C}^{N/2}$$

and it is equivalent to the following convex optimization problem [11] which we call Least Absolute Auto-Regressive (LAAR) Interpolator:

$$\begin{array}{ll} \text{variables} & \mathbf{e}, \mathbf{X}_u^{LAAR} \quad \text{complex} \\ \text{minimize} & \|\mathbf{e}\|_1 \\ \text{subject to} & \mathbf{e} = \mathbf{A}(\mathbf{K} \mathbf{X}_k + \mathbf{U} \mathbf{X}_u^{LAAR}) \end{array} \quad (11)$$

5. The vector obtained  $\mathbf{X}_u^{LAAR}$  is then incorporated into the DFT coefficients vector using the relations in (8):

$$\mathbf{X}_p^{OUT} = \mathbf{K} \mathbf{X}_k + \mathbf{U} \mathbf{X}_u^{LAAR}. \quad (12)$$

The other side of the DFT spectrum ( $[N/2 + 1, \dots, N]$ ) is also replaced with the new DFT coefficients exploiting the conjugate symmetry. We then anti-transformed in the time domain.

One question still standing is what to use as estimation of the AR vector  $\mathbf{a}$  in the MAP approach in (10). The problem considered is based on the following auto-regressive model, where the absolute value of the known and reliable DFT sample  $X_k(n)$  is written as a linear combination of  $P$  past samples:

$$X_k(n) = \sum_{p=1}^P a_p^* X_k(n-p) + e(n). \quad (13)$$

Using the same error distribution as in (10), we find the AR coefficients by solving the following minimization problem:

$$\begin{aligned} & \text{variables} && \mathbf{e}, \mathbf{a} \quad \text{complex} \\ & \text{minimize} && \|\mathbf{e}\|_1 \\ & \text{subject to} && \mathbf{e} = \mathbf{X}_k - \mathbf{M}_k \mathbf{a}^* \end{aligned} \quad (14)$$

where:

$$\mathbf{X}_k = \begin{bmatrix} X_k(N_1) \\ \vdots \\ X_k(N_2) \end{bmatrix}, \mathbf{M}_k = \begin{bmatrix} X_k(N_1-1) & \cdots & X_k(N_1-P) \\ \vdots & & \vdots \\ X_k(N_2-1) & \cdots & X_k(N_2-P) \end{bmatrix}$$

The starting and ending points  $N_1$  and  $N_2$  can be chosen in various ways assuming that  $X_k(n) = 0$  for  $n < 1$  and  $n > N_k$ . We set  $N_1 = K + 1$  and  $N_2 = N_k$ . If the error minimization criterion would have been the least squares (LS) it would have lead us to the covariance method [12]. It is important to notice that we are only considering the DFT coefficients that we labeled as known and reliable  $\{X_k(n)\}$ . The order  $P$  can be chosen on a frame-by-frame basis using a model-order selection approach [12] or can be chosen as fixed.

This scheme is incorporated into an analysis-modification-synthesis (AMS) framework and the output signal is resynthesized in a overlap-add fashion applying a gain correction due to the initial windowing. In our experimental analysis we used an hamming window of  $N = 256$  and before the next DFT computation the window is shifted by  $R = 128$  samples. The algorithmic delay introduced is 12.8 ms at 20 KHz. In this case we have found  $N_k$  is usually between 90 and 110 and the order  $K$  is usually between 20 and 30. We chose a fixed order AR analysis with  $P = 30$ . Examples are shown in figure. 1 and 2.

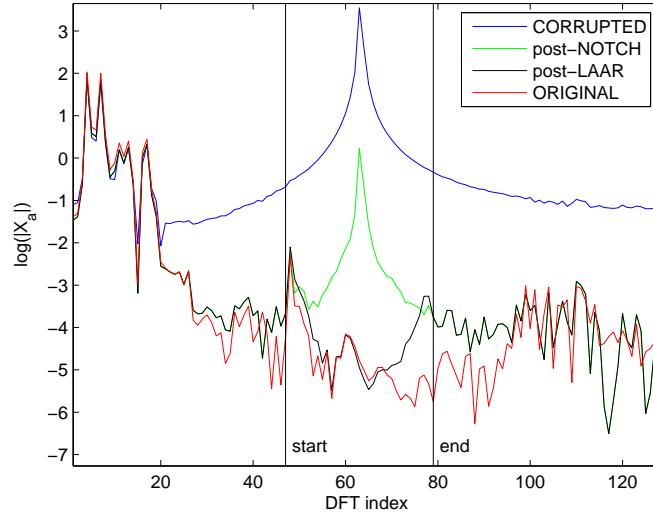


Figure 1: *Example of functioning of the LAAR interpolation with  $\omega_d = 63$  and  $B_d = 22$  (notch filter working with  $p = 0.7$ ). The figure shows the log-magnitude domain of a speech signal.*

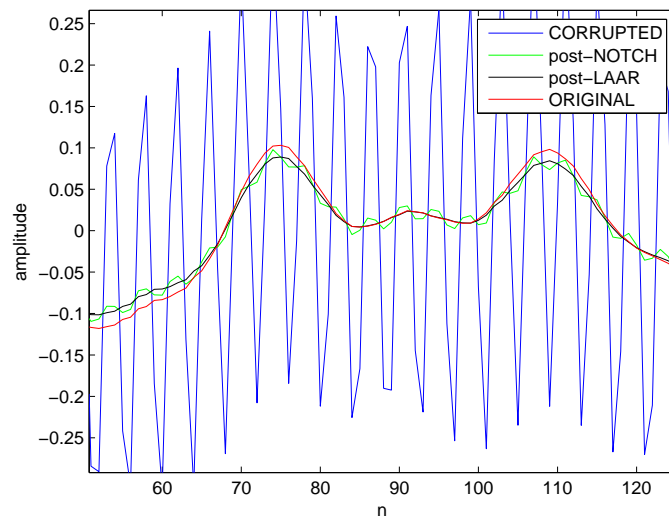


Figure 2: *Example of functioning of the LAAR interpolation in time domain.*

## References

- [1] A. Spriet, G. Rombouts, M. Moonen and J. Wouters, “Adaptive feedback cancellation in hearing aids”, *Journal of the Franklin Institute*, vol. 343, pp. 545–573, 2006.
- [2] N. F. Thornhill, B. Huang and H. Zhang, “Detection of multiple oscillations in control loops”, *Journal of Process Control*, vol. 13, no. 1, pp. 91–100, 2003.
- [3] N. Westerlund, M. Dahl and N. Grbic, “Detection and attenuation of feedback induced howling in hearing aids using subband zero-crossing measures”, *Proceedings of the Eighth International Symposium on Signal Processing and Its Applications*, vol. 2, 2005.
- [4] J. Wei, L. Du, Z. Chen and F. Yin, “A new algorithm for howling detection”, *Proceedings of the 2003 International Symposium on Circuits and Systems*, vol.4, 2003.
- [5] J. A. Maxwell and P .M. Zurek, “Reducing acoustic feedback in hearing aids”, *IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 4, pp. 304–313, Jul. 1995.
- [6] S. Haykin, *Adaptive filter theory*, Prentice-Hall, 1996.
- [7] S. Bittanti, M. Campi and S. M. Savaresi, “Unbiased estimation of a sinusoid in colored noise via adapted notch filters”, *Proc. Automatica*, vol. 33, no. 2, pp. 209–215, 1997.
- [8] S. V. Vaseghi and P. J. W. Rayner, “Detection and suppression of impulsive noise in speech communication systems”, *Proc. IEEE*, vol. 137, no. 1, pp. 38–46, Feb. 1990.
- [9] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration - a statistical model-based approach*, Springer-Verlag, 1998.
- [10] J.-H. Chang, “Complex Laplacian probability density function for noisy speech enhancement”, *IEICE Electron. Express*, vol. 4, no. 8, pp. 245–250, Feb. 2007.
- [11] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.



- [12] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Pearson Prentice Hall, 2005.