preprint no. 1266 (M-5)

AUTOMATIC DIGITAL AUDIO PROCESSOR - ADAP

By

J.E. Paul
Electronics Research Center
Rockwell International
Anaheim, California

presented at the 58th Convention November 4-7, 1977 New York

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

AUTOMATIC DIGITAL AUDIO PROCESSOR - ADAP

By

J. E. Paul
Electronics Research Center
Rockwell International
Anaheim, California

Abstract

Predictive deconvolution is a digital signal decorrelation technique for reducing the adverse effects of additive and convolutional noise on audio signals. This procedure makes use of a Wiener predictor to estimate the correlated noise components in an audio signal and to subtract these components from that signal. The resulting signal is then shaped to a desired spectrum.

An adaptive predictor further enhances this process by reducing undesired noise effects which are only short-term stationary. A real time 150$^{\underline{th}}$ order adaptive predictive deconvolver, the Automatic Digital Audio Processor (ADAP), has been developed and is being used successfully in a variety of digital audio processing applications. Several examples of noise cancellation on audio recordings using the ADAP will be presented.

Introduction

This paper discusses two digital audio processing techniques for cancelling noise on audio signals. The first technique, adaptive predictive deconvolution, uses an adaptive linear predictor to estimate and cancel (time) correlated noise components on an audio signal. The second technique, adaptive filtering, employs two audio signal inputs, the first having the desired audio signal along with noise and the second sensing principally the noise. The second, noise signal is adaptively filtered and subtracted from the first signal cancelling noise components common to the two inputs.

A stand-alone digital signal processor has been developed to carry out these two noise cancellation techniques. This Automatic Digital Audio Processor, ADAP, carrys out these processes in real time with up to a 256th order digital filter.

Audio Noise Characterization

Two linear noise processes, additive noise and convolutional noise, are generally introduced into an audio signal either at the audio microphone, acoustically, or during the recording/transmission of the signal electrically. Additive noise is defined as noise from external sources which is summed with the desired audio signal. Examples of additive noise include background environmental noise and electrically coupled 60 Hz buzz. Seldomly is additive noise of a total random nature, but it often contains significant correlated components. Often the correlated noise components are only short-term stationary, however.

Convolutional noise, on the other hand, does not result from external sources but is audio degradations resulting from convolving the signal (and its additive noise) with the acoustic resonances and transmission properties of the environment or the non-uniform transfer function of the audio transmission channel. Examples of convolutional noise include the reverberant effect of hard-walled rooms, the tinny effect of an early gramophone recording, and the muffled effects of unequalized telephone circuits. Generally, convolutional effects are very correlated with their source audio signal; however, their transfer functions may be time variant.

As might be expected, these two linear noise processes are not independent. As in the example of Figure 1, the desired voice signal, s, is augmented by additive noise from a muscial instrument, n_a, and is further degraded by the convolutional effects of the room acoustics, represented by the transfer function, H. The additive noise, in this example, is insufficient to mask the voice; however, once that noise, along with the voice signal, is convolved with reverberant room acoustics, the voice of the resulting signal, $r=H(s+n_a)$, is unintelligible.

Predictive Deconvolution

Predictive deconvolution is a powerful noise cancellation procedure which decorrelates the input signal, removing long-term correlated components. Figure 2 illustrates the process. The input audio signal is sampled yielding the sequence x_n, $n=1,2,\ldots$. A linear transversal predictor estimates the next signal sample, x_n, and subtracts this estimate, $\hat{x}_n$, yielding an output error sample

$$\varepsilon_n = x_n - \hat{x}_n \ .$$

Because a least-mean-square (LMS) estimator is used, the output error signal, to the degree of the predictor, is decorrelated in a long-term sense and has a flat power spectrum. A spectral shaper is thus introduced to restore the long-term power spectrum to the desired signal.

The predictor used in this procedure is a linear transversal digital filter whose output estimate, $\hat{x}_n$, is the weighted sum of past input signal samples, i.e.,

$$\hat{x}_n = \sum_{i=1}^{N} w_i x_{n-i} , \tag{1}$$

where w_i, $i=1,2,\ldots,N$ are the filter coefficients.

The order, N, required for the predictor depends upon the complexity of the time correlated signal components being removed. Simple tones added to voice signals can be removed with N being small, say less than 10. Convolutional noise, however, generally requires a much longer filter. Acoustic reverberations of rooms often require orders greater than 100 to adequately model room acoustics and longer echo paths.

The LMS estimator may be derived in the following manner. The output error sample from Figure 2 and equation (1) is

$$\varepsilon_n = x_n - \sum_{i=1}^{N} w_i x_{n-i} , \text{ or}$$

$$= \sum_{i=0}^{N} w_i x_{n-i} \qquad\qquad w_0 \triangleq -1 . \tag{2}$$

The expected value of the error signal squared is

$$E\{\varepsilon^2\} = E\left\{\left(\sum_{i=0}^{N} w_i x_{n-i}\right)^2\right\} . \tag{3}$$

The LMS solution is found by determining the set of coefficients, w_i, $i=1,2,\ldots,N$, which minimizes the value of equation (3). This is done taking the partial derivative of equation (3) with respect to each coefficient, observing the orthogonality principle (Papoulis, 1965), and setting each partial equal to zero, i.e.,

$$\frac{\partial}{\partial w_j} E\{\varepsilon^2\} = 2\, E\left\{\sum_{i=0}^{N} w_i x_{n-i}\, x_{n-j}\right\} = 0 \qquad j = 1,2,\ldots, N.$$

$$= \sum_{i=0}^{N} w_i\, R_x(j-i) = 0 , \tag{4}$$

where the autocorrelation function is defined as

$$R_x(j-i) = E\{x_{n-i} x_{n-j}\} .$$

The set of N equations, (4), is the coefficient solution for the causal Wiener predictor. By applying these coefficients to equation (2), it can be observed that the output error signal is decorrelated, i.e.,

$$R_\varepsilon(n) = 0 \qquad n = 1,2,\ldots, N \ . \tag{5}$$

From equation (5), it is obvious that long-term correlated components in the output signal are removed. Speech will pass through this process virtually unaffected, as it is only short-term stationary and has virtually no long-term correlated components[1].

Predictive deconvolution, though providing an optimal LMS noise cancelling process, has two major disadvantages when applied to real-world audio signals. First, the estimation of the a priori autocorrelation statistics and the solving of the N simultaneous equations, (4), require a significant amount of computing, especially when N is large. Secondly, the coefficient solutions are based on the assumption of long-term stationarity of the noise components, which often is not the case. These two disadvantages can be avoided and hardware realization more easily achieved by making the process adaptive.

Adaptive Predictive Deconvolution

Adaptive predictive deconvolution, illustrated in Figure 3, is achieved utilizing an adaptive transversal predictor to estimate and cancel the correlated signal components. An adaptive predictor (Widrow, et al, 1975) employs the output error signal to continuously adjust the predictor toward the LMS coefficient solution of equation (4). It can be shown that with a stationary input signal, the coefficients of this predictor may indeed converge to the Wiener solution.

The filter coefficients are derived by the adaptive processor using a steepest descent algorithm in which instantaneous (unbiased) estimates of the error surface gradient continuously correct the coefficients toward the LMS solution. The coefficients, initially set to any arbitrary value, are incremented each sample interval according to the following procedure:

$$w_n^{\,i+1} = w_n^{\,i} + \Delta w_n^{\,i} \ . \tag{5}$$

Here $w_n^{\,i}$ is the $n^{\underline{th}}$ predictor coefficient at the $i^{\underline{th}}$ time sample. That coefficient is incremented by $\Delta w_n^{\,i}$ which is computed in the following manner:

$$\Delta w_n^{\,i} = 2\mu\varepsilon x_{i-n} \ . \tag{6}$$

The parameter μ adjusts adaptation convergence rate. Widrow, et al (1975) point out that to assure convergence,

$$1/\lambda_{max} > \mu > o \ ,$$

[1] Actually, the non-flat long-term spectrum of speech introduces some coherencies; however, these may be easily restored with the spectral shaper of Figure 2.

where λ_{max} is the largest eigenvalue of the autocorrelation matrix, R. The effective convergence time, T, is related to μ and R, as

$$T = (N+1)/4\mu \text{ Trace } \{R\} .$$

In practice, the autocorrelation matrix is usually unknown and the convergence times are adjusted empirically.

In addition to not requiring a priori signal statistics and to not solving equations (4), this adaptive approach allows time invariant noise sources to be cancelled, as long as these sources are correlated long enough for filter convergence. In practice, convergence times of a few hundred milliseconds are reasonable for N equal to a few hundred. This convergence property allows, for example, the tracking and cancellation of music on voice signals.

Adaptive Filtering

Adaptive filtering (Widrow, et al, 1975) illustrated in Figure 4, is similar to adaptive predictive deconvolution in structure. The principal differences is in the manner in which the estimate of the noise is derived. An adaptive filter is a two-input processor in which the second input is the reference noise input. The noise, n_2, of that input is correlated with the noise, n_1, of the primary input and is linearly filtered to form an estimate, $\hat{n}_1$, of the noise on the primary input. Since the desired signal, s, does not appear on the second input, an estimate of s cannot be produced by the linear filter.

The output error signal,

$$\varepsilon = s+n_1 - \hat{n}_1 ,$$

is minimized in a LMS sense by the adaptation process in the same manner as in the adaptive predictive deconvolver. The noise, n_1, is reduced without distortion to the desired signal s. Should noises n_1 and n_2 not be correlated with each other, no noise cancellation (or addition) is produced, since the filter coefficients converge to a zero solution.

The adaptive filter, by obtaining its reference external to the primary signal, does not require the noise be time correlated to be cancelled. Even random noise can be cancelled, as long as the two random noises, n_1 and n_2, are correlated with each other. Adaptive filtering will not, however, cancel noise which it cannot sense on its reference input. As a result, convolutional noise effects, such as room acoustics, are often difficult to cancel with adaptive filtering but are readily removed with predictive deconvolution.

The Automatic Digital Audio Processor - ADAP

The ADAP, illustrated in Figure 5, is a real-time digital signal processor developed to carry out both adaptive filtering and adaptive predictive deconvolution. Since its introduction in May, 1977, it has been successfully employed on over 50 law enforcement audio recordings for noise cancellation purposes. The ADAP has also been successfully employed in a number of other audio enhancement applications including aircraft cockpit noise cancellation, the restoration of early gramophone recordings, and noise cancellation on HF communications systems.

The ADAP is a self-contain instrument containing both analog preprocessing electronics and a high-speed digital pipeline processor. A full complement of panel controls allows the operator to specify processor parameters including convergence time, filter order (up to 256), bandwidth (up to 8 kHz), and a host of others.

Input audio to the ADAP is lowpass filtered by the antialiasing filters and sampled by the 12-bit sampling subsystem. Audio samples from the two audio inputs are then stored in two solid state memories. Transversal filtering is carried out multiplying the input audio samples by filter coefficients stored in the coefficient memory. Each product is accumulated completing the convolution operation of equation (1).

The output error signal is produced by differencing as in equation (2) and is analog converted by the digital-to-analog converter with 12-bit resolution. The adaptation processor then readjusts the filter coefficients as in equation (5).

The ADAP is very effective as a real-time signal processor. As an adaptive filter, it has cancelled over 20 dB of random background noise from voice signals. As an adaptive predictive deconvolver, it has attenuated pure tones nearly 50 dB on voice signals and substantially reduced the reverberant effects of room acoustics. The ADAP is adaptive and is, therefore, automatic. The operator does not "tune" the processor to cancel undesired noise effects but simply adjusts the parameters to within a broad range.

References

A. Papoulis, Probability, Random Variables, and Stochastic Processes, McGraw-Hill Book Company, New York (1965).

B. Widrow, J. Glover, J. McCool, J. Kounetz, C. Williams, R. Hearn, J. Zeidler, E. Dong, and R. Goodlin. "Adaptive Noise Cancelling: Principles and Applications," Proceedings of IEEE, Vol. 63, No. 12, Dec. 1975.

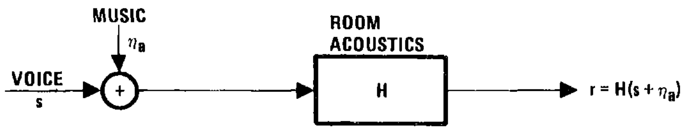

Figure 1. Example of Compound Room Effects

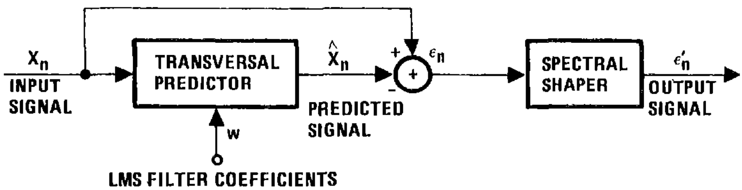

Figure 2. Predictive Deconvolution

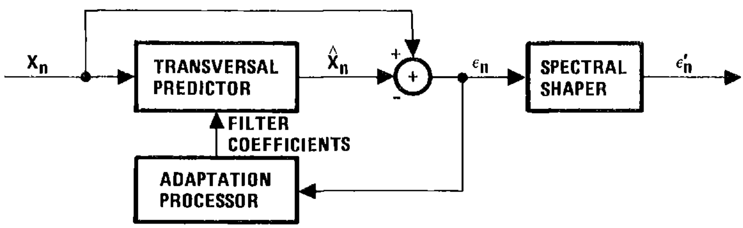

Figure 3. Adaptive Predictive Deconvolution

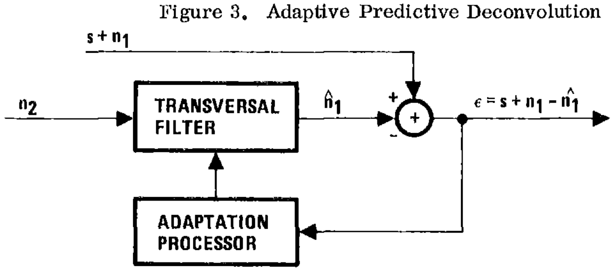

Figure 4. Adaptive Filtering

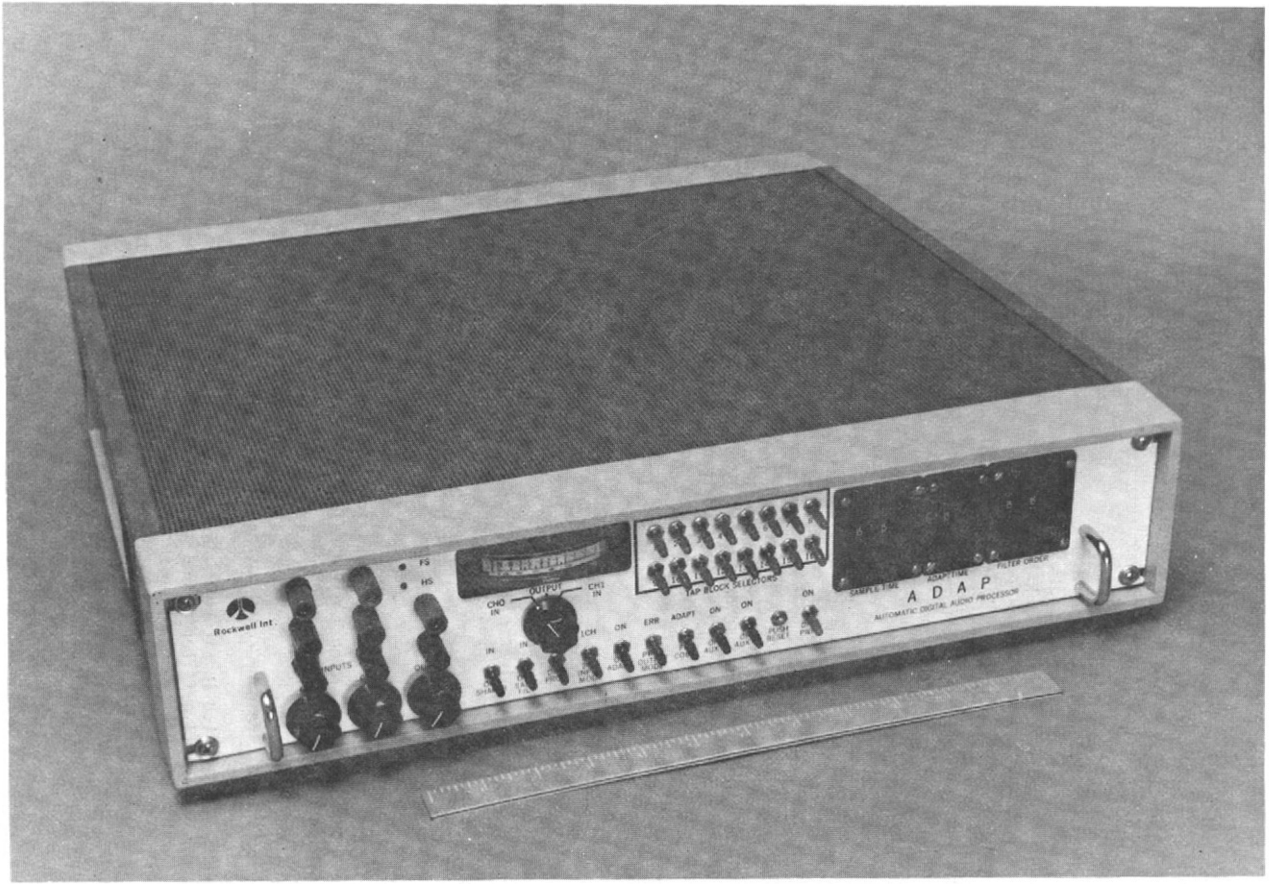

Figure 5. Real-Time Automatic Digital Audio Processor (ADAP)