DSP RESTORATION TECHNIQUES FOR AUDIO

James A. Moorer

Adobe Systems, Incorporated

ABSTRACT

The CD revolution of 1984 unleashed a massive campaign of rerelease of vintage analog recordings. To their collective horror, record labels found that many of the original master recordings had deteriorated. Often the only existing copy of a recording was a vinyl pressing. This rush to release beloved music in the new digital form spawned intense interest in digital methods of restoring these old recordings. A number of DSP techniques have been developed to give new life to these recordings. This paper is a survey of some of those methods.

Index Terms— Music, Signal Restoration, Signal Processing, Linear predictive coding, Wiener filtering

1. INTRODUCTION

When the compact disc (CD) was introduced in 1984, there was a scramble to create material for the new format. The five major record labels went back to the original masters of their catalog of material to bring all their existing recordings onto the new format. There was some great gnashing of teeth when it was discovered that the master recordings were often in terrible shape, or were missing. There soon developed significant interest in DSP techniques to repair some of the problems in these old recordings. It was found that in some cases, the audio could be restored to something close to the original quality.

When music is treated as a signal, we find it has certain properties that help simplify the processing. Specifically, music is largely piecewise autoregressive, or noise-like, or some combination of the two. I apologize for this terminology. When you call something "autoregressive", it implies that it can be generated by convolution over all time. By adding "piecewise", I am trying to say that the convolution kernel changes with time. Technically, it isn't really autoregressive any more. It is a bit like saying "piecewise stationary". This allows us to use second-order statistics to model the sound and reconstruct missing or damaged regions. Another important fact is that we are dealing with audio. The difference between audio and any other signal is that ultimately somebody listens to audio. Nobody listens to, say, radar signals. This means that the processing only has to deal with attributes of the signal that can be heard. Anything that can't be heard is irrelevant. This gives us important flexibility in the processing, but it also requires that we know something about how people hear.

2. HISTORICAL SUMMARY

One of the earliest restorations of musical sound by DSP was done by Stockham, *et al.* [1]. They applied digital techniques to deconvolve the effects of the horn on acoustic cylinder recordings of Enrico Caruso from 1907. Janssen, Veldhuis and Vries at Philips Laboratories pioneered a number of DSP techniques for restoring dropouts (missing samples) in CD playback [2]. The techniques they developed formed the basis for the click and pop removal used by later systems. In the late 80's, two commercial companies brought advanced DSP techniques to the market in time for the great CD rerelease wave. These were Cedar and Sonic Solutions. Cedar chose a hardware-based approach (that is, free-standing devices) versus the Sonic Solutions workstation-based system. 20 years later, Cedar still actively makes devices for studio use in sound restoration. Sonic Solutions licensed the NoNoise® technology to SonicStudioHD which still sells digital audio workstation software systems. These are the industrial-strength professional commercial solutions. I should point out that just about every digital audio workstation now has some capability for signal restoration. These are all based on the techniques I will describe below. I personally am a co-founder of Sonic Solutions and wrote the NoNoise® package. Much of the implementation detail remains proprietary, but I can discus some of the public-domain fundamentals of the restoration techniques.

3. TYPES OF DAMAGE

The kinds of problems encountered in restoring vintage recordings can roughly be grouped into three areas (although some would argue there are an endless number of ways a recording can be damaged). These are broad-band noise (hiss), impulsive noise (clicks and pops), and parasitics (hum, turntable rumble). Each requires its own technique.

4. BROAD-BAND NOISE (HISS)

In general, there is no way to bring a signal back when it is below the noise floor. If it happens to last a long time, then conceivably the noise could be averaged out, leaving only the desired signal. With music, it is generally not possible to do this. Part of the interest of music is that it is always changing, so we must deal with it the way it is. The best we can hope for is to try to decide which parts of the spectrum are music and which parts are noise, then attenuate the portions that are noise.

All techniques for broad-band noise reduction can be derived from the Wiener-Hopf filter [3]:

$$\frac{\Phi_{ss}}{\Phi_{nn} + \Phi_{ss}}$$

This includes such popular techniques as spectral subtraction [4]. The techniques differ in their interpretation of the terms of the Wiener filter and how they are calculated. In general, we assume that the noise floor has a roughly constant spectrum. We can then get a sample of the noise spectrum in a place where the desired signal is very quiet or not present. If such a place doesn't exist in the audio sample, then sometimes it can be estimated frequency band by frequency band by looking for minima in the spectrum over time.

Once we have the noise spectrum, we have to come up with a formula for determining the gain of each frequency band of the spectrum. One formula might be as follows [5]:

$$g_i = (1-\alpha)\left[\frac{s_i}{s_i + \lambda_i}\right]^r + \alpha$$

In this formula, g_i represents the gain of frequency band *i*, s_i is the signal in frequency band *i*, λ_i is the noise floor in frequency band *i* (or some quantity derived from the noise floor), *r* represents the "abruptness" of the gain curve and α represents the amount of attenuation to be applied. Note that if α is zero and *r* is unity, this corresponds exactly to the Wiener-Hopf filter. The parameters *r* and α are adjusted by the operator for a pleasing sound. The idea is to construct a filter by computing the gain in each frequency band. If a band is to be passed without change, the gain will be set to 1.0.

If you just apply this formula as written, you quickly discover that there are a number of problems. The gain will audibly "pump" – that is, changes in gain from one block to the next will be audible. This formula may attenuate high frequencies or transient signals more than desired. Frequencies below 1500 Hz present additional problems. The gains of those lower bands are generally coerced to unity (although care must be taken to avoid large discontinuities between the coerced gains and uncoerced gains). In general, the implementation of a broad-band noise-reduction system is an art and requires a great deal of hand-tuning. The bands must be "opened" (that is, gains rise to unity) quickly as a sound begins and must be "closed" (gains drift towards α) somewhat more slowly. This is where a good understanding of loudness [6] and psychoacoustics [7] becomes important.

5. IMPULSIVE NOISE (CLICKS)

Vinyl records deteriorate both with age and with playing. As few as five playings of a vinyl disc is enough to erode most of the high frequencies. Moreover, the old 78 RPM records were made without a hardener component in the vinyl. This causes the discs to become very brittle as they age. Some of the best Louis Armstrong recordings we have of his early material are 78 RPM records that were transferred to magnetic tape in the late 1940's. Many of the original 78 RPM records, even ones that have never been played, are unplayable now.

This sad state of affairs, plus the fact that the original tapes are often lost (if they ever existed), leads us to try to save these recordings from vinyl discs despite the problems. Most of the difficulty is in missing vinyl – places where the vinyl has broken off or been scratched. Although this is a somewhat arbitrary distinction, we generally divide the problem into two categories: "clicks", which tend to be isolated, and "crackle", which tends to be a continuous sound, sometimes described as sounding like a rain forest. A click can be as large as several milliseconds, but is often spaced more than 25 milliseconds from adjacent clicks. Crackle is generally sub-millisecond but can be spaced by just a millisecond or two (or less).

These defects represent missing material. We have to come up with ways to locate these defects and then to synthesize material to fill the gaps. The usual way this is done is to assume that the signal is autoregressive in the region of the defect [2, 8]. Once we have estimated an autoregressive model, we can locate the defects since they are not autoregressive. If we simply filter each region of the sound with the inverse filter, we can expect each defect to provoke a large output from the inverse filter, since it does not match the model. There is a chicken-and-egg problem in that we have to compute the model before we detect the defects. To do this, we have to assume that the defects represent a relatively small proportion of the total number of samples (they do) such that the power spectrum of the music dominates the estimation process (it generally does). After the defects are located, a refined estimate of the model can then be made without using samples known to be defective.

Once we have the autoregressive model of the good samples, a simple linear least-squares process is sufficient to solve for the missing samples. To see how this works, let us write the equations for autoregressive modeling. We start with the linear filtering of some input samples, x_k, by a linear filter, a_j, to produce an output, ε_k:

$$\varepsilon_k = \sum_{j=0}^{N} a_j x_{k-j}$$

Note that $a_0 \equiv 1$. We now square both sides of this equation and sum over a number of samples:

$$\varepsilon_k^2 = \left(\sum_{j=0}^{N} a_j x_{k-j} \right)^2$$

To derive the coefficients, we differentiate this formula with respect to one of the unknown coefficients. The result is a set of linear equations in the unknown coefficients that is straightforward to solve. Solving for the unknown coefficients in the presence of missing or unknown input samples requires a slight reformulation that is beyond the scope of this paper. It proceeds by noting that a sequence with missing samples can be considered to be a set of separate sequences that are taken from the same ensemble: *i.e.*, from the same distribution. To solve for unknown samples, we use exactly the same procedure, but now the coefficients are known and some number of the data samples are unknown. We differentiate the formula with respect to the unknown samples, then rearrange the formula so that it forms a set of linear equations in the unknown samples. I might warn that this system of linear equations is not well-conditioned and requires some care to solve in a stable manner. Figure 1 shows the result of applying this technique to two clicks in a piece of audio (specifically, a recording of Patsy Cline singing "Crazy").

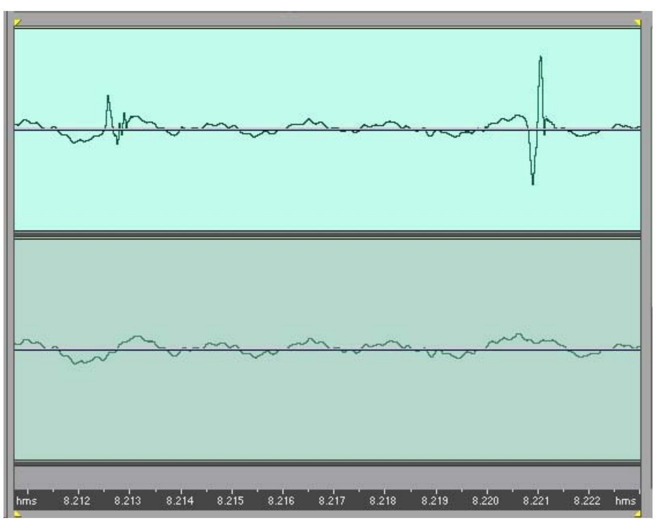

Figure 1: Signal with clicks (upper) and same signal with clicks repaired by assuming the surrounding signal is autoregressive.

The same technique can be used to solve for both clicks and crackle. There are some differences in implementation that are appropriate. For clicks, we generally use a larger basis for computing the coefficients than we do for crackle. It should be clear that the error in the interpolation increases rapidly with the number of consecutive unknown samples. The condition of the solution equations also worsens with increasing gap size.

6. HUM

One form of parasitic is hum from the power lines. One would think this would be just a simple sinusoid which could be easily removed out with some kind of notch filter, but one would be wrong. The more problematic form of hum involves hundreds of harmonics. This sounds like a buzz at 60 (or 50) Hz. I am not sure of the mechanism of production of these parasitics, but I can say that they are quite common in audio and video field recordings where the electrical environment is not well-controlled.

The first method I tried to eliminate this form of noise was to simply use banks of hundreds of notch filters. This produced an odd audible artifact consisting of a slap echo. That is, there was a noticeable echo of the original sound, delayed by 16.67 (or 20) milliseconds. This is simply due to the fact that a filter consisting of a single delayed impulse is a comb filter, which has a zero of transmission at integral multiples of the frequency represented by the delay. Even if a subset of the zeros are included, the resulting impulse response still resembles a delay. Some artistry and experimentation is required to detune and widen the notches by just enough to reduce the hum but still not degrade the desired musical sound. Note that there are numerical problems with implementing a composite filter consisting of hundreds of second-order sections. Such a filter require special care in implementation.

7. DIRECTIONS FOR FUTURE WORK

Despite these advances, there are still a number of improvements that could be made, and problems that we haven't attempted to solve to date. For instance, the 78 RPM records only record frequencies up to about 7.5 kHz. Is there any way to restore the higher frequencies? They are physically not present, so they would have to be created from scratch. At this time, there is no way to do that. We can imagine many ways that might work, but I find it hard to imagine something that works consistently and without the reconstructed harmonics popping in and out as the signal gets louder and softer.

Another issue is that the way I have described to replace missing samples assumes that the residual of the implied inverse filter is zero. This assumption is flatly false. When interpolating over a relative wide area, significant "droop" can be noticed in the synthesized waveform. To some extent, this can be fixed by using some amount of white noise as a synthetic residual signal. This fixes the case of droop, but only when the "true" residual, which can't be directly observed, is white. In cases such as brass instruments (trombone, trumpet) or human voice, the residual is not white noise, but is more impulsive. At this time, there is no suitable method for choosing the correct residual function.

With hum removal, the problem with using hundreds of notch filters is fundamental and unavoidable. There needs to be a different technique entirely for this kind of parasitic.

There are any number of other problems that I have not mentioned. For instance, overload distortion is quite common. In the case of hard limiting, such as is common in digital systems, the autoregressive assumption can be used to put the tops (and bottoms) back on the waveforms. When the overload is caused by analog saturation, it could possibly be modeled by some kind of polynomial mapping or Volterra expansion.

A further problem is wind noise. In this case, the noise is not stationary. The technique I described above for hiss removal can be used, but what happens is that the recovered signal pumps badly as the noise level changes. When the noise level is high, there are no high frequencies in the recovered signal. They come back as the noise level drops. This makes a very annoying change in the tone quality in the recovered signal.

8. SUMMARY

I have described three kinds of problems with historic recordings that have been repaired, more or less, using DSP techniques. I can confidently make the claim that if you have a CD containing music that is more than 30 years old, there is about an 85% chance that it has been restored using one or more of these techniques. They are so common-place today that any serious DAW (digital-audio workstation) generally comes with some kind of implementation of one or more of these tools.

9. REFERENCES

[1] Thomas G. Stockham Jr., Thomas M. Cannon, Robert B. Ingebretsen "Blind Deconvolution Through Digital Signal Processing" Proceedings of the IEEE, Volume 63, Number 4, April 1975, pp678-693

[2] A.J. Janssen, R. Veldhuis, L.B. Vries "Adaptive Interpolation of Discrete-Time Signals That Can Be Modelled as Autoregressive Processes" IEEE Transactions on Acoustics, Speech and Signal Processing, Volume ASSP-34, Number 2, pp317-330, June, 1986.

[3] Norbert Wiener *Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications* 1930 (reissued in 1964 by MIT Press)

[4] Steven F. Boll "Supression of Acoustic Noise in Speech Using Spectral Subtraction" IEEE Transactions on Acoustics, Speech and Signal Processing, Volume ASSP-27, Number 2, pp113-120, April 1979.

[5] James Moorer, Mark Berger "Linear-Phase Bandsplitting: Theory and Applications" Presented at the 76th Convention of the Audio Engineering Society, October 8-11, 1984, New York, Preprint 2132 (session A-1).

[6] Eberhard Zwicker, Bertram Scharf "A Model of Loudness Summation" Psychological Review, Volume 72, Number 1, February, 1965, pp3-26

[7] Hugo Fastl, Eberhard Zwicker *Psychoacoustics: Facts and Models* Springer-Verlag, 2006, 462p.

[8] Simon Godsill, Peter Rayner *Digital Audio Restoration: A Statistical Model Based Approach*, Springer-Verlag, 1998, 328p.