AUDIO SIGNAL EXTRAPOLATION - THEORY AND APPLICATIONS

Ismo Kauppinen and Kari Roth

University of Turku
Department of Physics
FIN-20014 Turku, Finland
iska@utu.fi, kari.roth@utu.fi

ABSTRACT

A method for extrapolating discrete audio signals is described. The theory of extrapolation is studied and some applications are presented and demonstrated. The extrapolation method is fast and capable of extrapolating several thousand samples of CD-quality audio signals. The extrapolation is applied in practice to enhance the spectral resolution in short-time fast Fourier transform based methods. It is also applied to eliminate impulsive noise bursts and to recover missing signal sections.

1. INTRODUCTION

The extrapolation of a discrete signal section means calculation of new previously unknown signal samples extending the given signal. The conventional method for data series extrapolation is to fit the data into an assumed functional form and to find the best fitted parameters of the analytical equation. When the analytical equation is obtained, the data values outside the range of the given data series can be calculated using the analytical equation. With audio signal the exact functional form of the signal is hardly ever known. In the proposed method the extrapolated data values are calculated directly from the given data and no functional form nor other information about the data structure is required, which makes the method well suited for audio signal extrapolation. Other methods for discrete signal extrapolation can be found in e.g. [1]-[6].

The discrete audio signal extrapolation method presented in this paper was introduced in paper [7]. In this paper the theory of the extrapolation is derived with an alternative approach starting from the conventional linear prediction. Some applications for the audio signal extrapolation are presented to demonstrate the effectiveness of the presented method.

The remaining text is organized as follows. In Section 2, the linear prediction is briefly reviewed. In Section 3, the forward extrapolation equation is presented. In Section 4, theoretical aspects of extrapolating audio signals are studied. Applications for the extrapolation are presented in Section 5, and conclusions are drawn in Section 6.

2. LINEAR PREDICTION

In conventional Linear Prediction (LP) the nth signal sample x_n is approximated as a combination of p previous samples and computed using an finite impulse response (FIR) filter:

$$\tilde{x}_n = \sum_{i=1}^{p} a_i x_{n-i}, \tag{1}$$

where a_i are the prediction coefficients and p is the model order. The error between the predicted sample $\tilde{x}_n$ and the actual sample x_n is given by

$$e_n = x_n - \tilde{x}_n = x_n - \sum_{i=1}^{p} a_i x_{n-i} \tag{2}$$

and is called the residual. The unknown filter coefficients a_i are found by minimizing the square of the prediction error within a given signal section. The prediction filter is presented in the z-domain as

$$P(z) = \sum_{i=1}^{p} a_i z^{-i}, \tag{3}$$

and the filter $A(z) = 1 - P(z)$ is used to calculate the residual e_n from the signal x_n. If the residual e_n obtained by filtering x_n with the filter $A(z)$ is used as input to a infinite impulse response (IIR) filter given by

$$H(z) = \frac{1}{1 - P(z)}, \tag{4}$$

the signal x_n will be ideally recovered. Thus, the filter $H(z)$ is called the synthesis filter and the filter $A(z) = 1/H(z)$ the inverse filter. The frequency response of the all pole-filter $H(z)$ is used to model matches to the local maxima of the signal spectrum $X(f)$.

This gives rise to Linear Predictive Coding (LPC) which is an effective method for compression of speech and audio signals. In the compression applications relatively low model orders are used ($p = 8 - 12$ for speech signals [8] and $p = 30 - 100$ for audio signals [9]). This is because the aim in LPC is to model the strongest local maxima in the signal spectrum.

3. EXTRAPOLATION

3.1. Extrapolation as convolution

The extrapolation of a finite length signal vector $\mathbf{x} = [x_1, x_2, x_3, ..., x_N]$ means calculation of new previously unknown samples $[x_{N+1}, x_{N+2}...]$ (forward) or $[..., x_{-2}, x_{-1}, x_0]$ (backward) to the discrete signal vector $\mathbf{x}$. If we assume that there exists a set of prediction filter coefficients $\mathbf{h} = [h_1, ..., h_M]$ that would linearly predict any sample in a given signal perfectly (from M previous samples) resulting in zero prediction error ($e_n = 0$), Eq. (2) reduces to the form

$$x_n = \sum_{i=1}^{M} h_i x_{n-i}. \tag{5}$$

If we have at least M known samples in the given signal vector $\mathbf{x}$ (i.e. $N \geq M$), we can generate the first forward extrapolated sample x_{N+1} by the above equation resulting in a prolonged signal vector $[x_1, x_2, x_3, ..., x_N, x_{N+1}]$. Now the last M samples of this prolonged signal can be used to generate the second forward extrapolated sample x_{N+2} using Eq. (5) again. By successively using this procedure we can generate an unlimited amount of new extrapolated samples to the given signal **x**.

The extrapolation Eq. (5) can be rewritten in general convolution form

$$x_n = h_n * x_n = \sum_{i=-\infty}^{\infty} h_i x_{n-i}, \tag{6}$$

where $h_i = 0$ when $i > M$ and $i < 1$. This means that the impulse response for extrapolation must be causal and it must also satisfy the important condition $h_0 = 0$.

3.2. Impulse response and transfer function

To obtain some theoretical information about which type of signals can be extrapolated, it is essential to examine the impulse response in the frequency domain (i.e. the transfer function).

If we rewrite the convolution Eq. (6) in continuous form where a continuous signal $x(t)$ is convolved by an impulse response $h(t)$ so that it is not changed

$$x(t) = h(t) * x(t), \tag{7}$$

it can be written in frequency domain using the convolution theorem of the Fourier transforms as:

$$X(f) = H(f)X(f), \tag{8}$$

where the transfer function must obey the condition

$$H(f) = \begin{cases} 1 \text{ and real}, & X(f) \neq 0 \\ \text{arbitrary}, & X(f) = 0. \end{cases} \tag{9}$$

A trivial solution for a transfer function that satisfies the condition of Eq. (9) is a real constant value of unity at all f. The impulse response in this case is Dirac's delta function $h(t) = \delta(t)$. However, this cannot be used for extrapolation since according to Eq. (6) the impulse response must be causal and zero at $t = 0$. According to these requirements the transfer function $H(f)$ is complex valued and it is a Hermitian function i.e. $H(f) = H^*(-f)$. Furthermore, the real and imaginary parts of $H(f)$ are a Hilbert transform pair. An analytical function can only be forced to a certain value in discrete points, therefore the spectrum $X(f)$ of the infinitely long signal can consist only of sharp lines (i.e. Dirac's delta functions).

The above considerations can be summarized giving the requirements for the signals that can be extrapolated: *If the functional form of the given signal section has a theoretical spectrum consisting only of infinitely sharp lines, the signal section can be extrapolated perfectly using a finite length impulse response.*[1] An example of such a function is the cosine-function whose spectrum is given by

$$\mathcal{F}\{\cos(2\pi f_0 t)\} = \frac{1}{2}\delta(f - f_0) + \frac{1}{2}\delta(f + f_0). \tag{10}$$

where $\mathcal{F}\{\}$ denotes continuous Fourier transform.

[1] It is assumed here that the signal section has at least the same length than the required impulse response for perfect extrapolation.

4. EXTRAPOLATION OF AUDIO SIGNALS

A short stationary audio signal section can be mathematically approximated by a sum of cosine waves with the frequencies f_i and phases ϕ_i multiplied by an amplitude envelope functions $A_i(t)$ given by

$$x(t) = \sum_i A_i(t) \cos(2\pi f_i t + \phi_i), \qquad f_i \geq 0. \tag{11}$$

The amplitude envelope functions are slowly varying for relatively stationary signal sections, but for transient sounds, the amplitude envelopes have short rise and decay times and have a strong contribution to the waveform. The cosine function can be further decomposed into superposition of complex waves called 'phasors' according to Euler's formula

$$\cos \omega t = \frac{e^{i\omega t} + e^{-i\omega t}}{2}, \tag{12}$$

where ω is the angular frequency. The spectrum of a single phasor is Dirac's delta function and one single impulse response coefficient is required to extrapolate a phasor

$$e^{i\omega n\Delta t} = h_1 e^{i\omega(n-1)\Delta t}, \qquad \text{where} \qquad h_1 = e^{i\omega\Delta t}. \tag{13}$$

Two real valued coefficients are required to extrapolate a cosine wave, which is a sum of two phasors:

$$\begin{aligned} \cos(\omega n\Delta t) &= h_1 \frac{e^{i\omega(n-1)\Delta t} + e^{-i\omega(n-1)\Delta t}}{2} \\ &+ h_2 \frac{e^{i\omega(n-2)\Delta t} + e^{-i\omega(n-2)\Delta t}}{2}, \end{aligned} \tag{14}$$

where the impulse response coefficients are $h_1 = 2\cos(\omega\Delta t)$ and $h_2 = -1$.

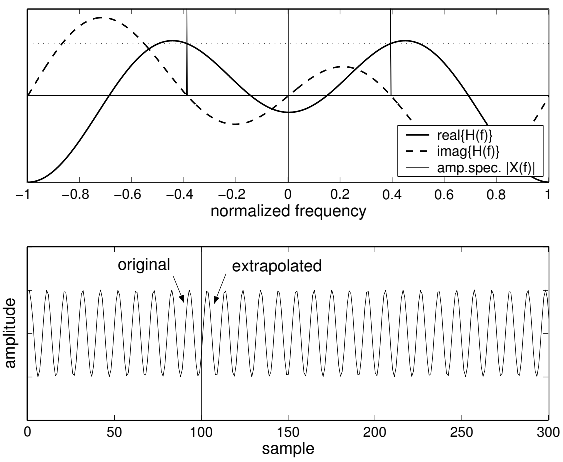

Figure 1: *Real and imaginary parts of the transfer function, and the amplitude spectrum of a single cosine wave are presented in the upper graph. The extrapolation is demonstrated in the lower graph.*

The real and imaginary parts of the transfer function for extrapolation of a single cosine wave are demonstrated in Fig. 1. In the upper graph, the real and imaginary part of the transfer function

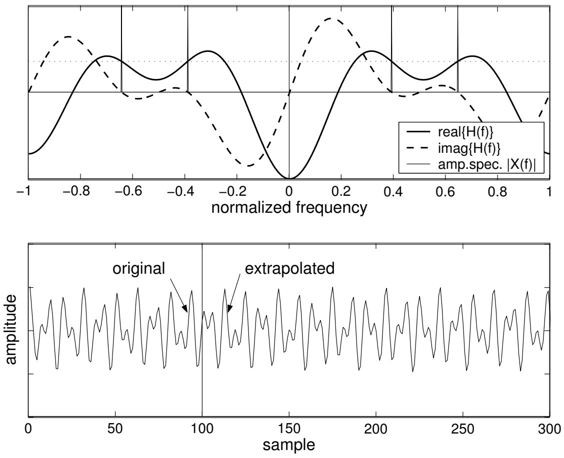

Figure 2: *The transfer function for the sum of two cosine waves in the upper graph. The extrapolation is demonstrated in the lower graph.*

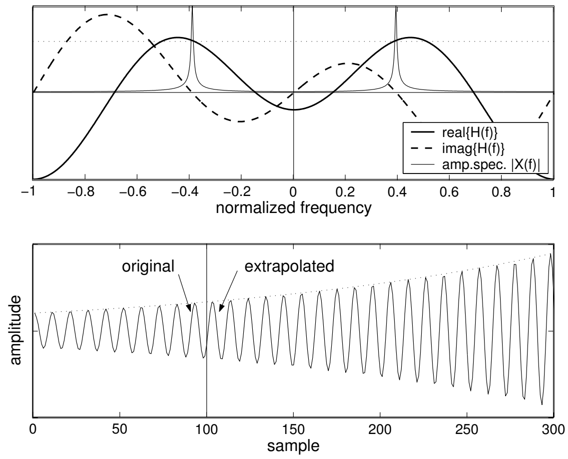

Figure 3: *The exponentially growing cosine wave can be perfectly extrapolated using the same amount of impulse response coefficients as for the same cosine wave with constant amplitude envelope.*

are plotted along with the amplitude spectrum of the (truncated) cosine wave. The dotted line indicates the value 1. The real part of the transfer function has the value of one and the imaginary part has the value of zero exactly at the locations of the positive and negative frequencies. In the lower graph, a section of the cosine wave is extrapolated perfectly using an impulse response with two coefficients.

The sum of two cosine waves with different frequencies (and constant amplitude envelopes) requires four impulse response coefficients for perfect extrapolation. This is demonstrated in Fig. 2 where the value of the transfer function is exactly one and real at the locations of the four spectral peaks in the amplitude spectrum of the signal.

In the case of a time-varying cosine wave, the perfect extrapolation is possible only if the amplitude envelope is a predictable function alone e.g. exponential or polynomial. For perfect extrapolation of a cosine wave with a non-constant amplitude envelope, a longer impulse response is usually required. An interesting case is the exponential amplitude envelope. An exponent function requires only one impulse response coefficient for perfect extrapolation and a cosine function multiplied by an exponential amplitude envelope requires only two coefficients which is the same amount than with a constant amplitude. The extrapolation of a cosine wave with exponential amplitude envelope is demonstrated in Fig. 3. (The broadening of the spectral lines results from the truncation of the signal.)

4.1. Model order

The number of impulse response coefficients required to perfectly extrapolate each time varying cosine wave in the audio signal can be observed by decomposing the cosine wave in the exponential form

$$x(t) = A(t)\cos(\omega t) = \frac{A(t)}{2}e^{i\omega t} + \frac{A(t)}{2}e^{-i\omega t}. \qquad (15)$$

If m is the number of coefficients required to perfectly extrapolate the amplitude envelope function $A(t)$, also $A(t)$ multiplied by an exponent function can be perfectly extrapolated with m coefficients. Each component in the right hand side sum in Eq. (15) requires m coefficients which means that a cosine wave with a time varying amplitude envelope requires $2m$ coefficients for perfect extrapolation. A polynomial function requires $q + 1$ impulse response coefficients for perfect extrapolation, where q is the order of the polynomial. Therefore, e.g. a cosine wave with a third degree polynomial decay requires 8 coefficients for perfect extrapolation.

General audio signals contain a large amount of frequencies and the time varying nature of the frequencies require higher model order than constant amplitude envelope. This implies a very large model order for good extrapolation results. Experiments with music signals in [7] and [10] suggest that an appropriate general number for the impulse response coefficients is 1000.

4.2. Calculation of the impulse response coefficients

A straightforward method for calculating the M impulse response coefficients is to apply Eq. (5) to a known section of the signal with N samples generating a group of M equations. This group of equations is given in matrix form as:

$$\mathbf{Xh} = \mathbf{x}, \qquad (16)$$

where $\mathbf{h} = [h_1, h_2, ..., h_M]^T$, $\mathbf{x} = [x_{M+1}, x_{M+2}, ..., x_{2M}]^T$, and $2M = N$. The matrix $\mathbf{X}$ is composed of shifted signal samples

$$\mathbf{X} = \begin{pmatrix} x_M & x_{M-1} & x_{M-2} & \dots & x_1 \\ x_{M+1} & x_M & x_{M-1} & \dots & x_2 \\ \vdots & \vdots & \vdots & & \vdots \\ x_{2M-1} & x_{2M-2} & x_{2M-3} & \dots & x_M \end{pmatrix}. \qquad (17)$$

However, the exact analytical solution for $\mathbf{h}$ exists only for noiseless theoretical signals. For measured noisy signals, such as audio signals, an iterative approach should be used.

The coefficients can be found by calculating the prediction error coefficients $\mathbf{a} = [1, a_1, ..., a_p]$ by LPC analysis and con-

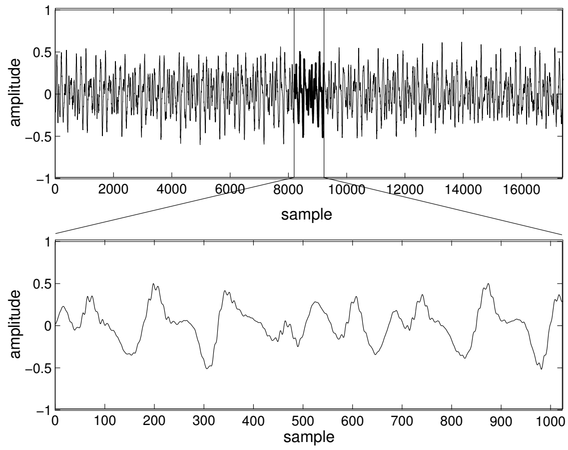

Figure 4: *The original 1024 samples signal frame (bottom) containing playing of an electrical guitar is extrapolated 7680 samples in both directions (top).*

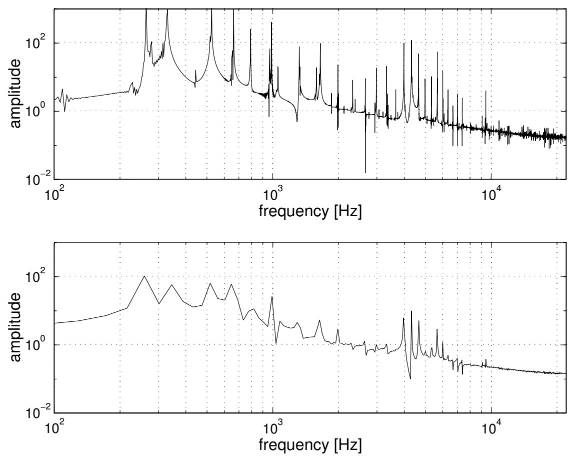

Figure 5: *The high resolution spectrum (top) with the resolution enhanced by a factor of 16 compared to the original spectrum (bottom).*

vert them into impulse response coefficients using the relation: $\mathbf{h} = [h_1, ..., h_M] = [-a_1, -a_2, ..., -a_p]$.

Several methods exist for finding the prediction error coefficients. However, in practical experiments the Burg method [11] [12][13] gives good and stable results for the purpose of audio signal extrapolation.

4.3. Information distribution

When extrapolated samples are generated by convolution some information is drawn from the known signal and the rest of the information comes from the impulse response coefficients. The impulse response coefficients bear mainly the information about the frequencies of the sinusoids and their amplitude envelopes. The amplitude and phase information of the extrapolated sinusoids comes from the known signal.

4.4. An IIR filter implementation of the extrapolation

A general infinite impulse response (IIR) filter is defined by equation [14]

$$\sum_{i=0}^{p} a_i\, y_{n-i} = \sum_{i=0}^{q} b_i\, x_{n-i}, \tag{18}$$

where a_i and b_i are the filter coefficients, $p+1$ and $q+1$ are the lengths of the filter coefficient vectors, x_n are the input samples and y_n are the output samples of the IIR filter. Assuming zero input to the filter (i.e. $x_n = 0,\ \forall n$) the output y_n can be solved from Eq. (18):

$$y_n = -\frac{1}{a_0} \sum_{i=1}^{p} a_i\, y_{n-i}. \tag{19}$$

This equation is essentially the same as Eq. (5). Therefore, the extrapolation can be implemented as an IIR filter with coefficients

$$\begin{aligned} a_0 &= 1 \\ a_i &= -h_i, \qquad i > 0 \\ b_i &= \text{arbitrary}, \end{aligned} \tag{20}$$

where h_i are the impulse response coefficients and the input is a vector of zeros. The most sensible choices for q and b_i are

$$q = 0,\ b_0 = 1 \tag{21}$$

leading into a difference equation given by

$$y_n = \sum_{i=1}^{p} h_i y_{n-i} + x_n. \tag{22}$$

The procedure for the extrapolation of W samples is:

- Calculate the impulse response coefficients $h_1, h_2, ..., h_M$.
- Initialize the filter with M past known samples just before the section to be extrapolated.
- Feed a zero vector of length W as an input to the filter.

The output of the filter will be the W extrapolated samples.

One possible implementation for the extrapolation using Matlab™ is given by

```
a = arburg(y,M);
Z = filtic(1,a,y(end-(0:(M-1))));
ye = filter(1,a,zeros(1,W),Z);
```

where `y` is a vector containing the known signal section, `M` is the length of the impulse response, and the vector `ye` contains the `W` extrapolated samples.

5. APPLICATIONS

5.1. Spectral resolution enhancement

In FFT based signal processing methods the signal is processed in short consecutive segments (frames) which are assumed to be stationary. To avoid sudden changes at the frame boundaries the frames can be overlapped. In the analysis stage FFT is applied to each signal frame to obtain the spectrum. This process has become known as the short-time Fourier transform (STFT). In the synthesis stage the frames are windowed and recombined by an

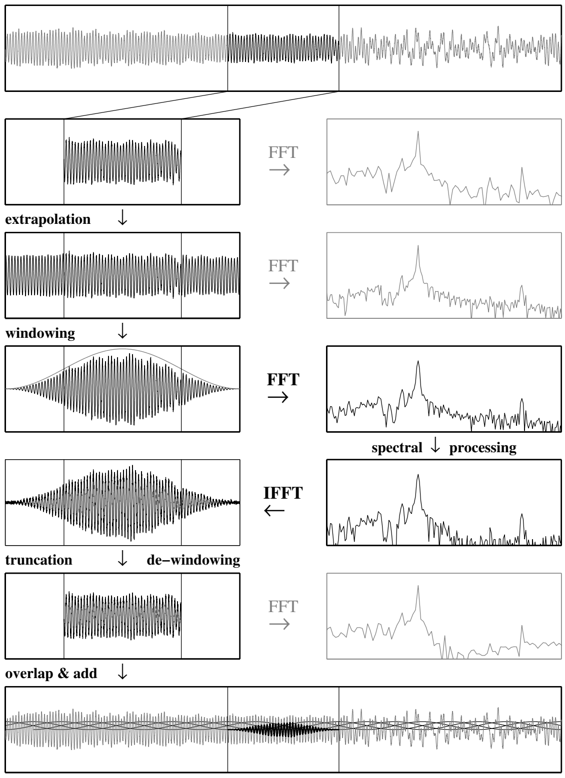

Figure 6: *The processing stages of one signal frame in a frame-by-frame FFT based signal processing system with resolution enhancement.*

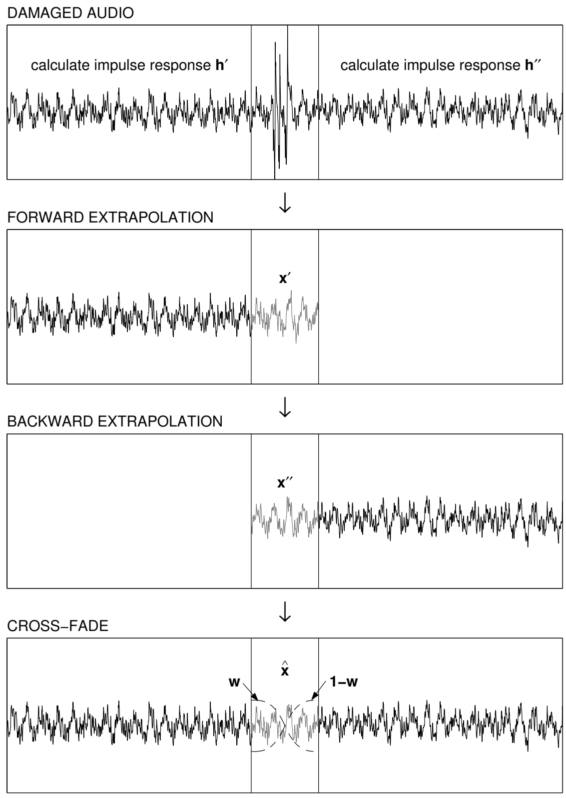

Figure 7: *Reconstruction of an audio signal section damaged by an impulsive noise burst. The signal contains Jazz music.*

overlap and add procedure. Audio signals contain fast transients and on the other hand long slowly varying sections i.e. the length of the short-time stationarity varies significantly as a function of time. To ensure the short-time stationarity assumption the signal should be processed in relatively short frames. On the other hand, high frequency resolution is required for high quality, and longer frames yield better frequency resolution. While the small frame size yields good time resolution, it also deteriorates the frequency resolution and vice versa. Traditionally, this phenomenon is known as the tradeoff between spectral and temporal resolution, which is a well-known problem in frame-by-frame FFT based audio signal processing techniques.

In the proposed resolution enhancement method, each stationary audio signal frame is prolonged by extrapolating in both directions. The prolonged signal frame is transformed into the spectral domain by using FFT, where the processing is applied with increased spectral resolution. The processed high resolution spectrum is inverse Fourier transformed back into the signal domain and the extrapolated (and processed) sections are discarded by truncation. This leaves a processed signal frame with the original frame size. However, frequency domain processing was applied with higher resolution yielding better accuracy.

The proposed spectral resolution enhancement is illustrated in Figs. 4 and 5 by using a real-life audio signal from an electric guitar. The size of the original frame is 1024 samples and it is modeled by calculating 800 impulse response coefficients using the Burg method. The signal section is extrapolated in both directions (7680 samples) resulting in a frame size of 16384 samples (see Fig. 4). The corresponding spectra of the original frame and the prolonged frame are presented in Fig. 5.

The processing stages of an arbitrarily selected signal frame in a frame-by-frame FFT based signal processing system with resolution enhancement is illustrated in Fig. 6. In this schematic figure the given signal frame is first prolonged by extrapolating in both directions. An analysis window is applied to the prolonged signal frame prior to FFT. (The half tone spectra merely illustrate the effect of the processing in frequency domain and do not represent actual processing.) After the desired spectral domain processing (e.g. spectral subtraction for noise reduction) the spectrum is inverse Fourier transformed and the analysis windowing is compensated. The extrapolated (and processed) sections are discarded by truncation resulting in a processed signal frame with the original frame size. Finally, overlap and add procedure is applied to construct the processed signal.

5.2. Elimination of impulsive noise

In audio signals impulsive type of degradation mainly originated from physical damages to the storage media is commonly encountered. A common example is a vinyl recording with scratches on the surface resulting in disturbing impulsive 'clicks' added to the original source.

Extrapolation can be successfully used to reconstruct an audio signal section damaged by an impulsive noise burst. The damaged signal section is replaced by a weighted combination of forward and backward extrapolated signals given by

$$\hat{x}_n = w_n x'_n + (1 - w_n)x''_n, \tag{23}$$

where x'_n and x''_n are the forward and backward extrapolated signal samples, respectively and w_n is the cross-fade function. The elimination of an impulsive disturbance by combining forward and backward extrapolation is demonstrated in Fig. 7.

5.3. Reconstruction of missing signal segment

The reconstruction of a data dropout is demonstrated in Fig. 8 by zeroing a 3000 samples section of an audio signal containing the playing of an acoustic guitar and reconstructing the missing samples by combining forward and backward extrapolated signals. The error of the extrapolation is visualized in the lowest graph in Fig. 8 by subtracting the reconstructed signal from the original signal.

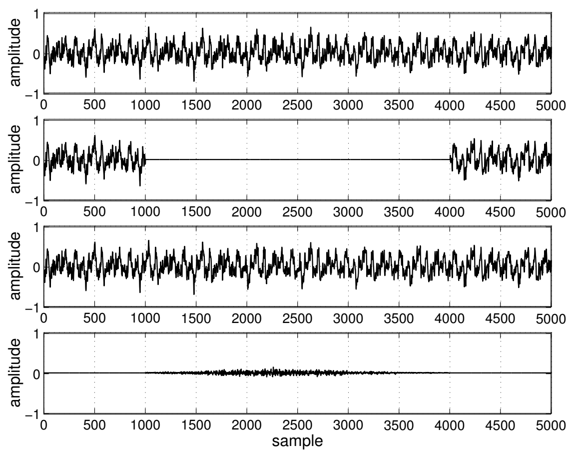

Figure 8: *A 3000 samples section of an audio signal (top) is zeroed (second graph) and reconstructed by combining forward and backward extrapolated signals (third graph). The error (original-reconstructed) is visualized in the lowest graph.*

6. CONCLUSIONS

A discrete signal extrapolation method is presented and theoretical observations are carried out. A mathematical model consisting of a sum of time varying cosine waves is assumed for audio signal and theory for extrapolating time varying cosine waves is discussed. Some applications for extrapolation in audio signal processing are presented including spectral resolution enhancement in frame-by-frame FFT based signal processing systems and reconstruction of damaged and missing audio signal sections by combining forward and backward extrapolation. The use of the proposed extrapolation method allows reconstruction of extremely long missing or damaged sections with non- or barely audible side-effects. Due to the low computational complexity, the extrapolation technique can be implemented into real-time audio restoration applications.

7. REFERENCES

[1] R. C. Maher, "A Method for Extrapolation of Missing Digital Audio Data," *J. Audio Eng. Soc.*, vol. 42, pp. 350-357, 1994.

[2] J. A. Cadzow, "An Extrapolation Procedure for Band-Limited Signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, pp. 4-12, 1979.

[3] S. D. Cabrera and T. W. Parks, "Extrapolation and Spectral Estimation with Iterative Weighted Norm Modification," *IEEE Trans. Signal Process.*, vol. 39, pp. 842-851, 1991.

[4] A. Papoulis, "A New Algorithm in Spectral Analysis and Band-Limited Extrapolation," *IEEE Trans. Circuits Sys.*, vol. CAS-22, no. 9, pp. 735-742, 1975.

[5] E. J. Diethorn and D. C. Munson, "A Linear, Time-Varying System Framework for Noniterative Discrete-Time Band-Limited Signal Extrapolation," *IEEE Trans. Signal Process.*, vol. 39, pp. 55-68, 1991.

[6] A. K. Jain and S. Ranganath, "Extrapolation Algorithms for Discrete Signals with Application in Spectral Estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-29, pp. 830-845, 1981.

[7] I. Kauppinen, J. Kauppinen, and P. Saarinen, "A Method for Long Extrapolation of Audio Signals," *J. Audio Eng. Soc.*, vol. 49, No. 12, pp. 1167-1180, 2001.

[8] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice Hall, Englewood Cliffs, NJ, 1984.

[9] S. Godsill, P. Rayner, and O. Cappe, *Digital Audio Restoration*, Springer-Verlag London, 1998.

[10] I. Kauppinen and J. Kauppinen, "Reconstruction method for missing or damaged long portions in audio signal," *J. Audio Eng. Soc.*, vol. 50, No. 7/8, 2002.

[11] F. Keiler, D. Arbif, and U. Zölzer, "Efficient linear prediction for digital audio effects," in *proc. DAFX-00*, Verona, Italy, pp. 19-24, 2000.

[12] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing*, 3rd edition, Prentice-Hall, New Jersey, 1996.

[13] S. Haykin, *Nonlinear Methods of Spectral Analysis*, Springer-Verlag, Berlin, 1983.

[14] S. Haykin, *Adaptive Filter Theory*, 3rd edition, Prentice-Hall, Upper Saddle River, NJ, 1996.