A Method for Extrapolation of Missing Digital Audio Data

3715 (B1-PM-7)

Robert C. Maher
University of Nebraska-Lincoln
Lincoln, NE 68588-0511, USA

Presented at the 95th Convention 1993 October 7–10 New York

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

A Method for Extrapolation of Missing Digital Audio Data

Robert C. Maher
Department of Electrical Engineering
and
Center for Communication and Information Science
University of Nebraska-Lincoln
209N WSEC, Lincoln, NE 68588-0511
Voice: (402) 472-2081 Fax: (402) 472-4732 Internet: rmaher@unl.edu

Abstract:

A method for extrapolating missing or corrupted samples in a digital audio data stream is presented. The method involves spectral extrapolation to synthesize an estimate of the missing material using a sinusoidal representation. The method takes advantage of the relatively slow variation in the time-variant spectral amplitude envelope in comparison with the relatively rapid oscillations of the time domain signal. Examples and applications are considered.

0. INTRODUCTION

In situations where a digital audio data stream contains segments of missing or corrupted data it is desirable to estimate the missing samples in order to reconstruct the signal with minimal audible defects. When the number of corrupted samples is small it may be possible to uniquely interpolate the missing material by making use of the bandlimited nature of the audio signal. Even linear interpolation or another simple method can audibly conceal the gap in many cases. However, gaps of hundreds or thousands of samples are long enough to affect many waveform periods, and in these cases it is necessary to establish a set of meaningful constraints to guide the extrapolation process.

In this paper an approach for extrapolating long segments of missing data is presented. The method assumes that uncorrupted signal segments precede and follow the missing data, and that the boundaries of the corrupted segment are known. The approach is to perform a time-variant spectral analysis on the uncorrupted signal samples both before and after the gap, then to interpolate across the gap using continuity constraints on the spectral amplitude and frequency information. Because this procedure requires knowledge of the gap boundaries it is most appropriate for use in off-line signal processing and restoration situations.

By performing the extrapolation operation in the amplitude-vs.-frequency-vs.-time domain it is possible to exploit the typically slow variation of the spectral amplitude envelope in comparison with the rapid signal oscillations in the time domain. Also, this representation is convenient for performing the required spectral extrapolation operations in a consistent and elegant manner.

The significance of this work is primarily in the specific methodology employed in obtaining an estimate for the missing signal material. The use of straightforward digital signal processing methods indicates that this approach is suitable for implementation in software on a wide range of non-realtime audio signal processing workstations. Thus, there are many situations where the process can be a beneficial addition to the audio engineering palette.

This paper continues with an overview of the general signal extrapolation problem. Next, the proposed analysis-synthesis strategy for audio signals is presented along with a description of the extrapolation procedure. Finally, several examples of the technique are presented, including a brief discussion of the results.

1. EXTRAPOLATION OF MISSING DIGITAL AUDIO SAMPLES

Extrapolation of unknown samples from known signal samples is an important task in many signal processing and signal estimation situations. Examples include the estimation of unknown meteorological or geophysical parameters based on limited physical measurements, reconstruction of tomographic or synthetic aperture images, and prediction of business and stock market cycles. In all applications the extrapolation procedure must incorporate some prior knowledge of the properties and expected behavior of the extrapolated signal.

A frequently encountered extrapolation situation involves *band-limited* signals, such as a Nyquist-sampled digital audio stream. If the number of samples to be extrapolated is small it is typically possible to obtain a unique solution using a standard band-limited interpolation approach [1, 2]. If, however, a large segment of signal must be extrapolated it is necessary to employ additional information and assumptions regarding the signal, such as minimum energy [3, 4, 5], spectral distribution weighting [6], parametric modeling [7, 8], and amplitude constraints [9].

The need for extrapolating missing samples in a digital audio data stream can occur whenever a segment of signal is lost due to a defective storage channel, a missing or delayed packet in a packet-switched network, a destructive editing or signal restoration procedure, etc. The principal difficulty in extrapolating audio signals is to determine which of the essentially infinite (although usually quantized in amplitude and discrete in time) possible sequences of samples is the "best" estimate of the missing material. Since the best estimate depends upon the perceptual transparency of the extrapolation it is difficult to express the optimum strategy as a simple least-squares minimization.

One way to view the extrapolation problem for audio signals is in terms of the time-variant spectrum of the known signal samples. Specifically, if the time-variant spectral envelope can be calculated for known signal segments both preceding and following the missing segment the extrapolation problem can be posed as a frequency-domain extrapolation problem. The primary advantage of this transformation is that the rapid oscillations of the time-domain audio waveform are avoided, while the relatively slow variation of the time-variant spectral envelope allows for an elegant analysis-synthesis extrapolation.

2. SINUSOIDAL ANALYSIS - SYNTHESIS FORMULATION

A sinusoidal time-variant spectral analysis/synthesis framework, published first by McAulay and Quatieri [10], has been found to be useful for representing speech, music, bioacoustical sounds, etc. [11-16]. The McAulay and Quatieri (or *MQ*) representation can be considered a generalization of simple Fourier series analysis to include time-variant spectra and possibly non-harmonic partials. In the implementation of the MQ analysis procedure used for the extrapolation problem considered in this paper the digitized input signal is divided into overlapping sections called *frames*. Each frame is multiplied ("windowed") by a lowpass window function to reduce spectral leakage, followed by calculation of a high resolution discrete Fourier transform (DFT) using a zero-padded Fast Fourier Transform (FFT) algorithm. The magnitude of the DFT is computed and all "peaks" in the magnitude spectrum are identified using interpolation and attributed to underlying sinusoidal components at those frequencies [11]. The amplitude, frequency, and phase corresponding to all of the spectral peaks in the frame are then calculated and recorded.

The DFT analysis and peak-picking process is repeated for each of the input frames and the spectral peak information (amplitude, frequency, and phase) is matched from frame to frame in order to follow changes in the input signal. The matching process results in connected sequences (or tracks) of peaks from frame to frame. The peak tracks are "born" and "die" as the spectral content of the signal varies with time. The peak matching process has the useful feature that at least first order amplitude, frequency, and phase continuity is assured. The MQ analysis system is depicted in **Figure 1**.

The input signal can be regenerated by an additive synthesis technique using the amplitude and frequency information obtained for each frame and a smoothly interpolated phase function, with cubic phase interpolation between blocks.

While the MQ process does not necessarily form a mathematically perfect analysis/synthesis system, the resynthesis results have been found to be excellent for many musical input signals [10]. Even signals such as broadband noise that are poorly described as a sum-of-sinusoids are synthesized with surprisingly good results for complex sonic textures [17].

3. EXTRAPOLATION USING MQ SPECTRAL ANALYSIS INFORMATION

The signal extrapolation problem can be visualized in terms of an MQ analysis sequence with one or more analysis frames missing. As an example, consider the simple amplitude modulated sinusoidal signal of **Figure 2**, where several cycles of the sinusoid are missing. The MQ analysis of the signal prior to the gap and after the gap is shown in **Figure 3**. The extrapolation technique proposed in this paper is to connect (extrapolate) the spectral tracks before and after the gap in order to resynthesize the missing portion of the signal. The extrapolation of the track makes use of the known magnitude, frequency, and phase of the track information available from the MQ analysis.

As mentioned previously, the MQ sinusoidal analysis uses the DFT of finite length windowed frames of the input signal, i.e.,

$$X(n,k) = \mathrm{DFT}\{\ w(n) \cdot s(n)\ \},$$

where n is the time index, k is the frequency index, w(n) is the low-pass window function and s(n) is the input signal. This expression can be converted to a frequency-domain convolution, viz.

$$X(n,k) = W(k) * S(k),$$

where W(k) and S(k) are the short-time transforms of w(n) and s(n), respectively. If we assume that the input signal spectrum S(k) is due to the presence of sinusoidal components (spectral impulses) the convolution operation results in the low-pass spectrum W(k) shifted and centered at the frequency of each input sinusoid. Observing X(n,k) as a function of n for a particular fixed value of $k=k_0$ reveals that this sequence can be interpreted as the output of a bandpass filter (shifted W(k)) centered at the frequency corresponding to the index $k=k_0$. Thus, the sequence of amplitude values comprising a track in the MQ analysis is bandlimited to twice the lowpass bandwidth of the window function. In practice this bandwidth is kept small enough to resolve individual partials of the input signal, indicating that extrapolation of the amplitude tracks can be very effective if the MQ frame rate (window overlap) is sufficient to obey the Nyquist theorem applied to sampling the short-time transform sequence [18].

3.1 Simple Linear Extrapolation

Consider the example signal depicted in **Figure 4**. The missing portion of the signal, 22.7 msec in this example (1000 samples at 44.1 kHz sample rate), is indicated by the segment with zero amplitude. The MQ analysis of the signal preceding and following the gap is shown in **Figure 5**. The extrapolation task at hand is to generate the missing MQ analysis frames across the gap, then to synthesize the time-domain signal. An elementary approach to the problem is to perform linear extrapolation on the amplitude and frequency information for each track.

The first step in the linear extrapolation process is to connect the tracks present at the beginning of the gap with the corresponding tracks at the end of the gap. If the gap is relatively brief compared to the rate of spectral change it is reasonable to match each track at the beginning of the gap to the track at the end of the gap with the smallest frequency difference. The next step is to calculate a linear trajectory (in amplitude and frequency) between each pair of matched tracks across the gap, as depicted in **Figure 6**. The measured phase information from the MQ analysis data is used to ensure time-domain waveform continuity at the gap boundaries. Finally, the extrapolated signal is synthesized using the calculated track information. The reconstructed signal for this example is shown in **Figure 7**.

3.2 Polynomial Extrapolation

If the spectrum of the signal is changing rapidly in the vicinity of the gap then a better match may be possible by observing the amplitude and frequency "trajectory" of each track in order to generate a smooth, polynomial extrapolation. A signal obtained from a recording of a soprano singer is shown in **Figure 8**, again with a 1000 sample gap. This signal contains both amplitude and frequency modulation, as shown in the MQ analysis of **Figure 9**.

Several possible polynomial extrapolation strategies are possible. A computationally simple method that has been found to work well in practice consists of the following steps:

Step 1: For each track present at the beginning of the gap, the frequency change between the two frames immediately preceding the gap is used to predict the terminal frequency of the track at the other end of the gap, under the assumption that the rate of frequency change remains relatively constant. The list of predicted terminal frequencies is sorted to avoid crossing tracks.

Step 2: The predicted terminal frequencies from step 1 are compared to the frequencies of the measured tracks at the end of the gap. A track at the beginning of the gap is linked to the track at the end of the gap with frequency closest to the predicted frequency.

Step 3: Using the track matching information from step 2, cubic polynomial extrapolation functions are generated for amplitude and frequency across the gap. The cubic functions are obtained simply by solving for the coefficients of the cubic function that passes through the two points preceding and two points following the gap. The required amplitude, frequency, and phase of the extrapolated gap frames is calculated from the cubic functions.

Step 4: Finally, the gap segment is synthesized from the extrapolated track information.

The cubic function extrapolation strategy applied to the MQ analysis data of Figure 10 is depicted in **Figure 10**. The synthesized time domain signal in the gap interval is shown in **Figure 11**.

4. CONCLUSION

This paper has presented a method for estimating missing or corrupted samples in a digital audio data stream. The method operates off-line by performing a spectral analysis of the audio signal both before and after the gap, extrapolating the spectral analysis information across the gap, and then synthesizing the missing audio samples. The procedure is based on the assumption that the spectral envelope of the audio signal changes more slowly than the time-domain features of the signal itself, thereby allowing a simple linear or low-order polynomial spectral extrapolation procedure.

In practice it has been found that the extrapolation procedure described in this paper is very effective for concealing gaps up to 30 msec in duration, although much longer gaps can be concealed if the signal spectrum remains relatively constant during the gap interval.

5. ACKNOWLEDGEMENTS

The work described in this paper was supported in part by the University of Nebraska Research Council and the Department of Electrical Engineering.

6. REFERENCES

[1] Schafer, R. W., and Rabiner, L. R., "A digital signal processing approach to interpolation," *Proc. IEEE*, **61**, no. 6, pp. 692-702, 1973.

[2] Crochiere, R. E., and Rabiner, L. R., *Multirate Digital Signal Processing,* Prentice-Hall, Englewood Cliffs, NJ, 1983.

[3] Papoulis, A., "A new algorithm in spectral analysis and bandlimited extrapolation," *IEEE Trans. Circuits and Systems,* **CAS-22**, no. 9, pp. 735-742, 1975.

[4] Cadzow, J. A., "An extrapolation procedure for bandlimited signals," *IEEE Trans. Acoust., Speech, Signal Processing,* **ASSP-27**, no. 1, pp. 4-12, 1979.

[5] Jain, A. K., and Ranganath, S., "Extrapolation algorithms for discrete signals with application in spectral estimation," *IEEE Trans. Acoust., Speech, Signal Processing,* **ASSP-29**, no. 4, pp. 830-845, 1981.

[6] Cabrera, S. D., and Parks, T. W., "Extrapolation and spectral estimation with iterative weighted norm modification," *IEEE Trans. Signal Processing,* **39**, no. 4, pp. 842-851, 1991.

[7] Janssen, A. J. M., Veldhuis, R. N. J., and Vries, L. B., "Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes," *IEEE Trans. Acoust., Speech, Signal Processing,* **ASSP-34**, no. 2, pp. 317-330, 1986.

[8] Diethorn, E. J., and Munson, D. C., Jr., "A linear, time-varying system framework for noniterative discrete-time bandlimited signal extrapolation," *IEEE Trans. Acoust., Speech, Signal Processing,* **ASSP-39**, no. 1, pp. 55-68, 1991.

[9] Abel, J. S., and Smith, J. O., "Restoring a clipped signal," *Proc. 1991 IEEE ICASSP,* paper D4a.7, pp. 1745-1748, 1991.

[10] McAulay, R.J. and Quatieri, T.F., "Speech Analysis/Synthesis Based on a Sinusoidal Representation," *IEEE Trans. Acoust., Speech, Signal Processing,* **ASSP-34**, no. 4, pp. 744-754, 1986.

[11] Smith, J. O., and Serra, X., "PARSHL: An Analysis/Synthesis Program for Non-harmonic Sounds Based on a Sinusoidal Representation," *Proc. 1987 Int. Computer Music Conf.,* pp. 290-297; San Francisco: Computer Music Assn., 1987.

[12] Maher, R. C., and Beauchamp, J. W., "An Investigation of Vocal Vibrato for Synthesis," *Applied Acoustics,* **30**, no. 2-3, pp. 219-245, 1990.

[13] Maher, R. C., "Evaluation of a Method for Separating Digitized Duet Signals," *J. Audio Eng. Soc.*, **38**, no. 12, pp. 956-979, 1990.

[14] Serra, X., "A System for Sound Analysis/Transformation/Synthesis Based on a Deterministic plus Stochastic Decomposition," Ph.D. Dissertation, Stanford University, Stanford, CA (also Dept. of Music Rep. STAN-M-58), 1989.

[15] Serra, X., and Smith, J.O., "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic plus Stochastic Decomposition," *Computer Music Journal*, **14**, no. 4, pp. 12-24, Winter, 1990.

[16] Quatieri, T.F., and McAulay, R.J., "Speech Transformations Based on a Sinusoidal Representation," *IEEE Trans. Acoust., Speech, Signal Processing*, **ASSP-34**, no. 6, pp. 1449-1464, 1986.

[17] Maher, R. C., "Sinewave additive synthesis revisited," *Proc. 91st Audio Eng. Soc. Conv.*, New York, NY, preprint 3128, October 1991.

[18] Portnoff, M. R., "Implementation of the digital phase vocoder using the fast Fourier transform," *IEEE Trans. Acoust., Speech, Signal Processing*, **ASSP-24**, no. 3, pp. 243-248, 1976.

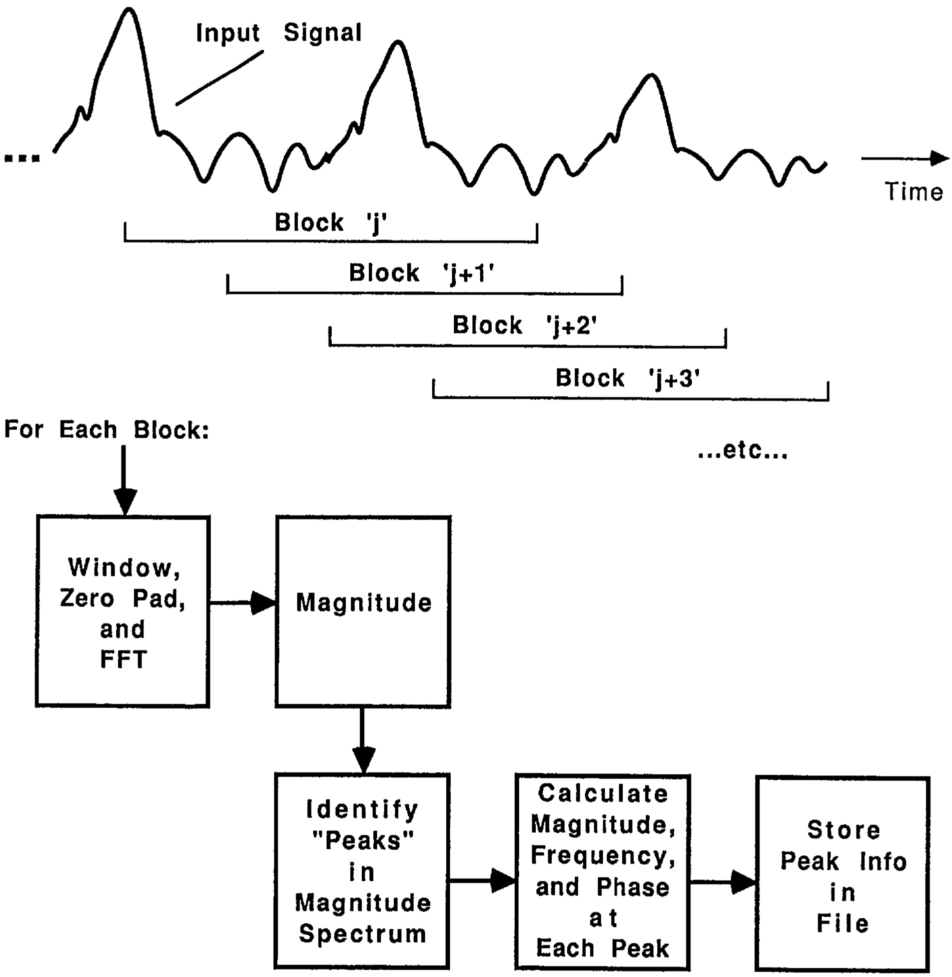

Figure 1: Block diagram of the McAulay-Quatieri (MQ) sinusoidal analysis scheme.

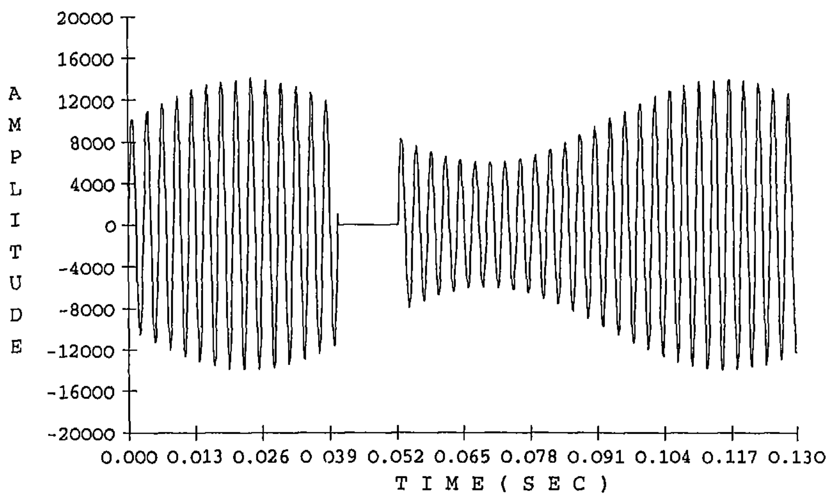

Figure 2: Sinusoidal signal with sinusoidal amplitude modulation and 11.6 msec gap (512 samples at 44.1 kHz sample rate).

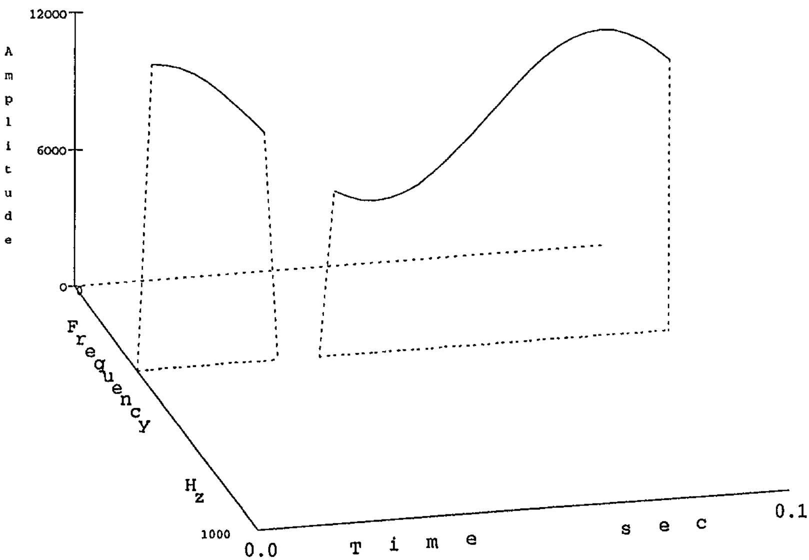

Figure 3: MQ analysis of the sinusoidally amplitude modulated sinusoidal signal of Figure 2 showing gap in the analysis data.

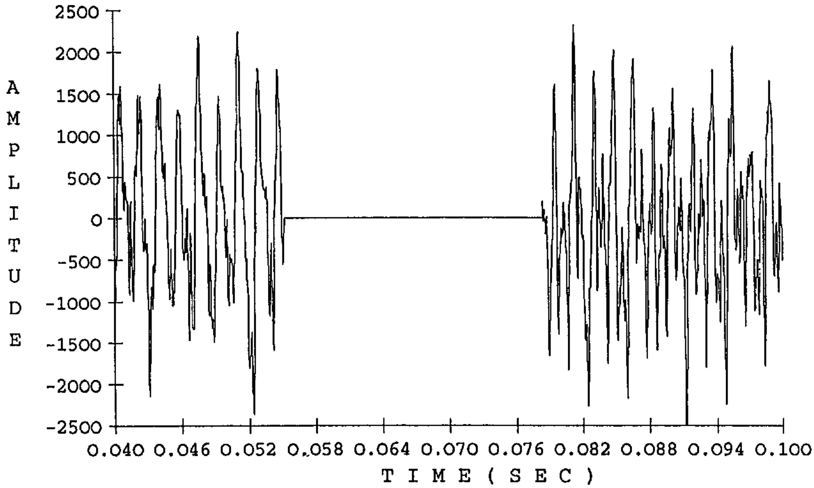

Figure 4: Example signal with 22.7 msec gap (1000 samples at 44.1 kHz sample rate).

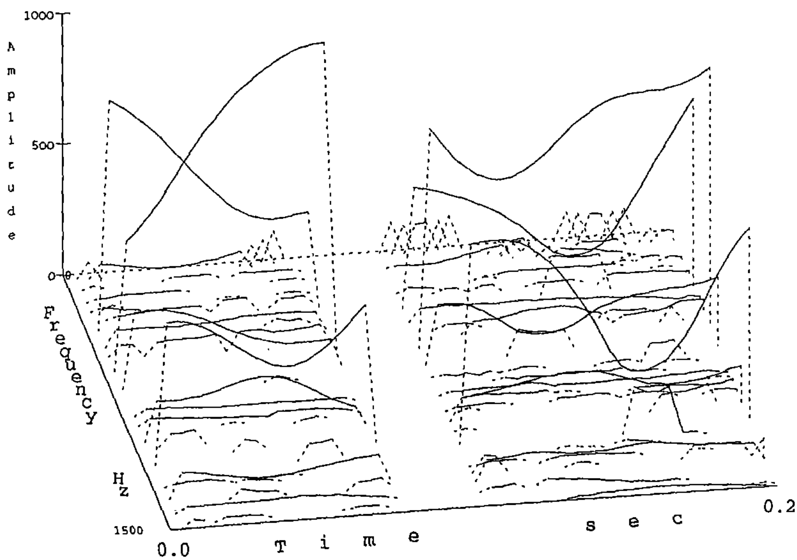

Figure 5: MQ analysis of the signal of Figure 4, with gap.

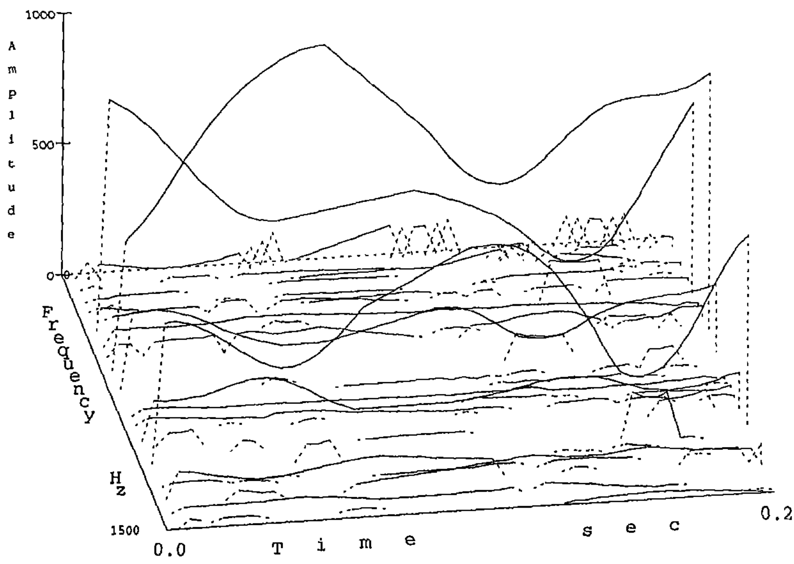

Figure 6: Nearest-neighbor match using linear extrapolation segments to fill the gap shown in Figure 5.

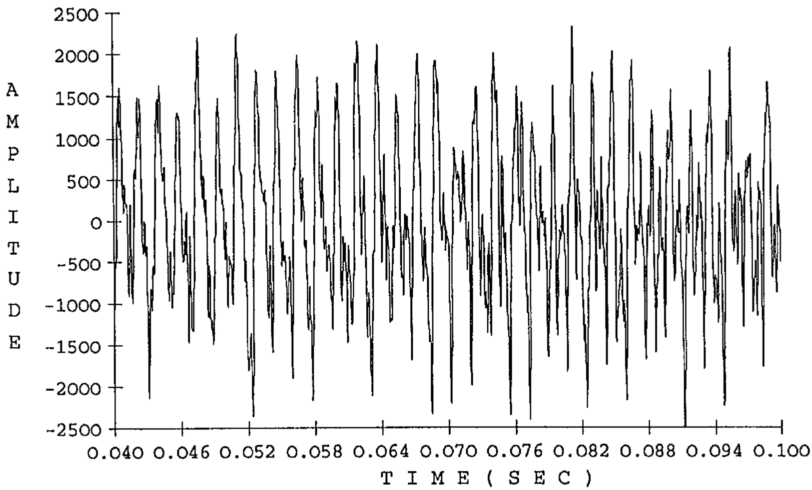

Figure 7: Reconstructed signal with gap fill resynthesized from the extrapolated MQ analysis data of Figure 6.

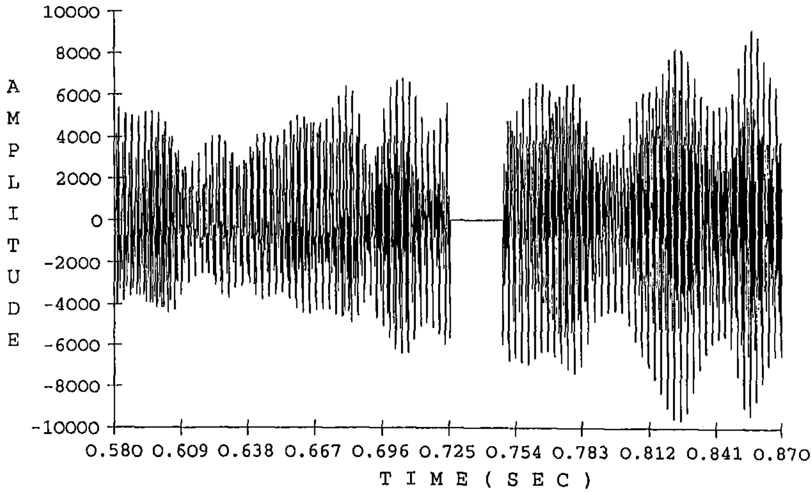

Figure 8: Example signal with 22.7 msec gap (1000 samples at 44.1 kHz sample rate).

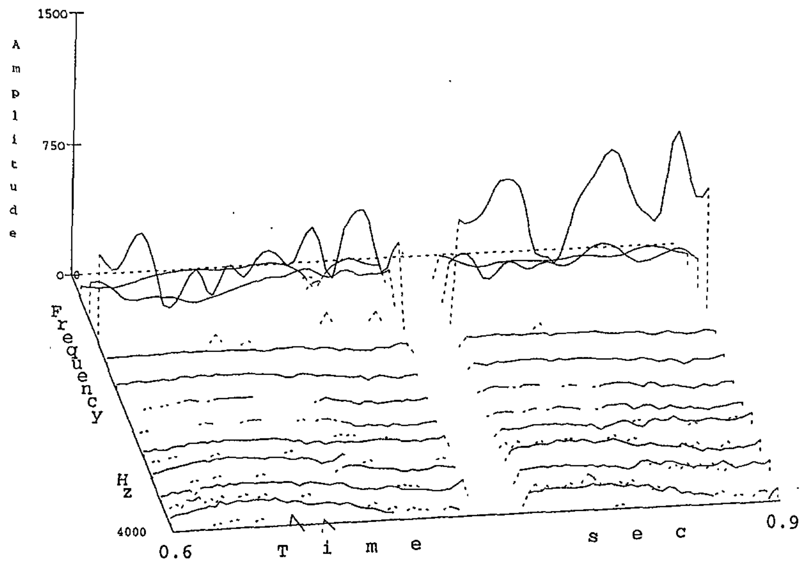

Figure 9: MQ analysis of the signal of Figure 8, with gap.

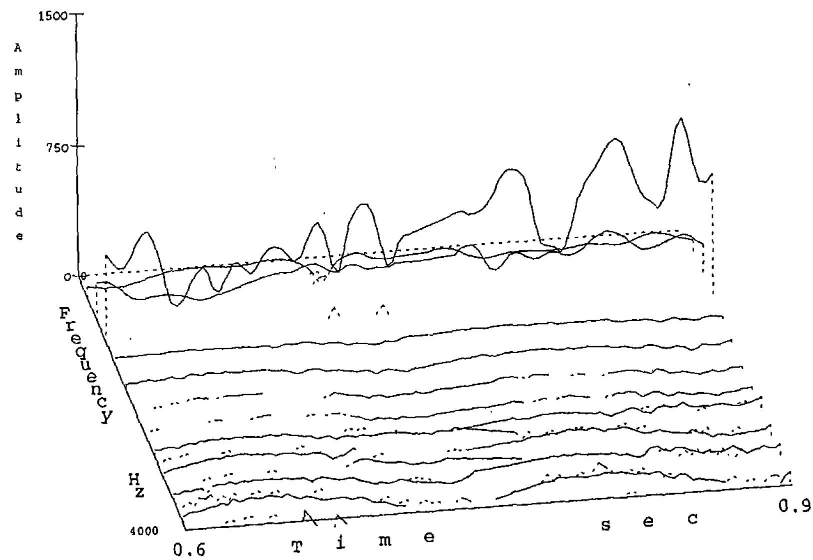

Figure 10: Cubic function polynomial extrapolation to fill the gap shown in Figure 9.

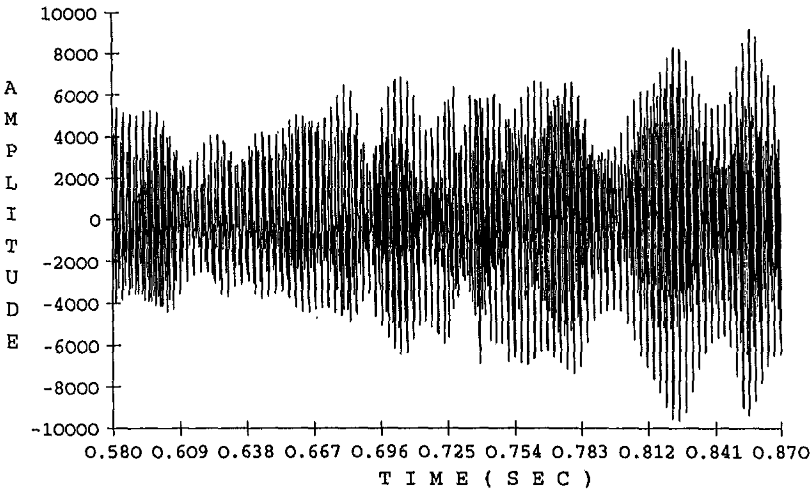

Figure 11: Reconstructed signal with gap fill resynthesized from the extrapolated MQ analysis data of Figure 10.