A Method for Extrapolation of Missing Digital Audio Data*

ROBERT C. MAHER, *AES Member*

Department of Electrical Engineering and Center for Communication and Information Science, University of Nebraska–Lincoln, Lincoln, NE 68588, USA

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, limitations, and applications are considered.

0 INTRODUCTION

Common methods for digital audio storage are generally very reliable, with intrinsic bit error rates below 10^{-10} in many cases. Situations do occur, however, where a digital audio data stream contains segments of missing or corrupted data because of mechanical failures, physical damage to the storage medium, or, whenever a segment of signal is simply unavailable due to a missing or delayed packet in a packet-switched network, a destructive editing or signal restoration procedure, and so on. When this happens it is desirable somehow to replace the invalid samples in such a manner as to reduce or eliminate any adverse audible effect. When the number of corrupted or missing samples is small, it may be possible to interpolate the missing material by making use of the band-limited nature of the audio signal. Even linear interpolation or another simple method can audibly conceal the gap in some cases. However, gaps of hundreds or thousands of samples are long enough to affect many waveform periods, and generally result in a thoroughly ill-posed mathematical reconstruction problem. In these cases it is necessary to establish a set of meaningful constraints to guide the extrapolation process.

In this engineering report a non-real-time approach for extrapolating long segments of missing data is presented. The method requires that uncorrupted signal segments precede and follow the missing data, and that the boundaries of the corrupted segment be known. A time-variant spectral analysis is performed on the uncorrupted signal samples both before and after the gap. Then the analysis data are extrapolated across the gap using continuity constraints on the spectral amplitude and frequency information. By performing the extrapolation operation in the amplitude–frequency–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.

It is important to note at the outset that the proposed technique does not solve all problems related to audio signal extrapolation. Rather, the significance of this work is primarily in the specific methodology used in obtaining an estimate for the missing signal material. Because this procedure requires knowledge of the gap boundaries, it is most appropriate for use in off-line signal processing and restoration situations. The use of straightforward digital signal processing methods indicates that this approach is suitable for implementation in software on a wide range of non-real-time audio signal processing workstations. Thus the process can be a beneficial addition to the audio enineering palette.

The narrative portion of this engineering report begins 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 are presented, including a discussion of the limitations of the technique.

1 EXTRAPOLATION OF MISSING DIGITAL AUDIO SAMPLES

Extrapolation of unknown signal samples from known samples is an important task in many signal processing

* Manuscript received 1993 August 18; revised 1993 December 27. Portions of this paper were presented at the 95th Convention of the Audio Engineering Society, New York, 1993 October 7–10.

and signal estimation situations: General 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. To be useful, the extrapolation procedure must incorporate some prior knowledge of the properties and the expected behavior of the extrapolated signal [1].

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 [2], [3]. 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 [4]–[6], spectral distribution weighting [7], parametric modeling [8], [9], and amplitude constraints [10].

The principal difficulty in extrapolating audio signals is to determine which of the practically 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 on 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, then the time-domain 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 [11], has been found to be useful for representing speech, music, bioacoustical sounds, and so on [12]–[17]. The McAulay–Quatieri (MQ) representation can be considered a generalization of simple Fourier series analysis to include time-variant spectra and possible nonharmonic partials. In the implementation of the MQ analysis procedure used for the extrapolation problem considered in this engineering report, the digitized input signal is divided into overlapping blocks called *frames*. Each frame is multiplied (windowed) by a low-pass 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 [12]. 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 amplitude, frequency, and phase continuity is assured. The MQ analysis system is depicted in Fig. 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 mathematical identity analysis–synthesis system, the resynthesis results have been found to be excellent for many musical input signals [11]. Even signals such as broad-band noise that are poorly described as a sum of sinusoids are synthesized with surprisingly good results for complex sonic textures [18].

3 EXTRAPOLATION USING MQ SPECTRAL ANALYSIS INFORMATION

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

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

where n is the time index, k the frequency index, $w(n)$ the low-pass window function, and $s(n)$ the input signal.

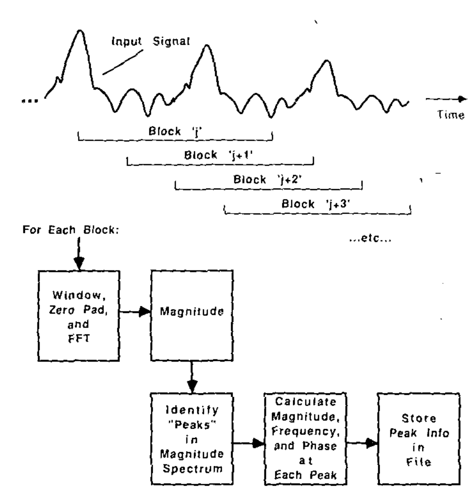


Fig. 1. Block diagram of McAulay–Quatieri sinusoidal analysis scheme.

This expression can be converted to a frequency-domain convolution,

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

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. Now 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 [the shifted low-pass $W(k)$] 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 band-limited to twice the low-pass 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 [19].

3.1 Simple Linear Extrapolation

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 Fig. 2, where 11.6 ms (512 samples at 44.1-kHz sample rate) of the signal are missing. The MQ analysis of the signal prior to the gap and after the gap is shown in Fig. 3. The extrapolation task at hand is first to generate the missing MQ analysis frames across the gap and 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. A similar concept has also been proposed independently by Hee-Yong Kim, a student at Rutgers University, in an unpublished term paper [20].

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 Fig. 4. 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 Fig. 5.

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 Fig. 6, here with a 1000-sample (22.7-ms) gap. This signal contains both amplitude and frequency modulation, as shown in

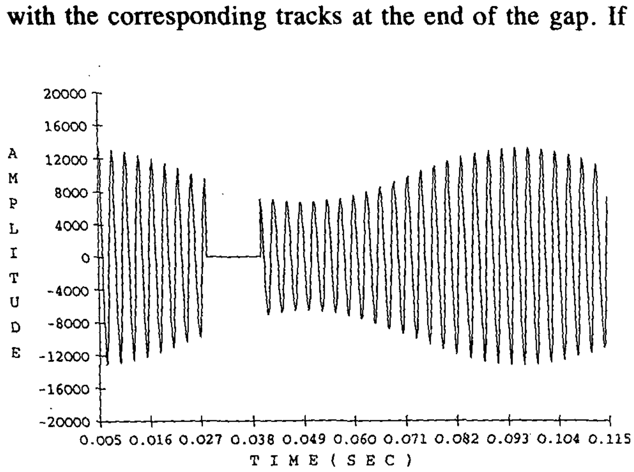


Fig. 2. Sinusoidally amplitude-modulated signal with 11.6-ms artificial gap (512 samples at 44.1-kHz sample rate).

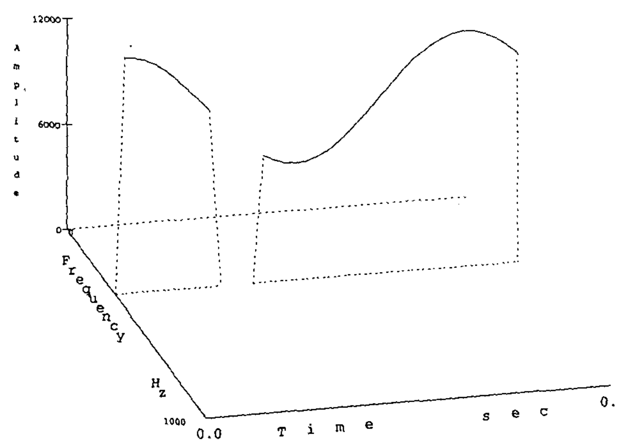


Fig. 3. MQ time-variant analysis of signal of Fig. 2. Note gap in analysis data corresponding to missing samples of input signal.

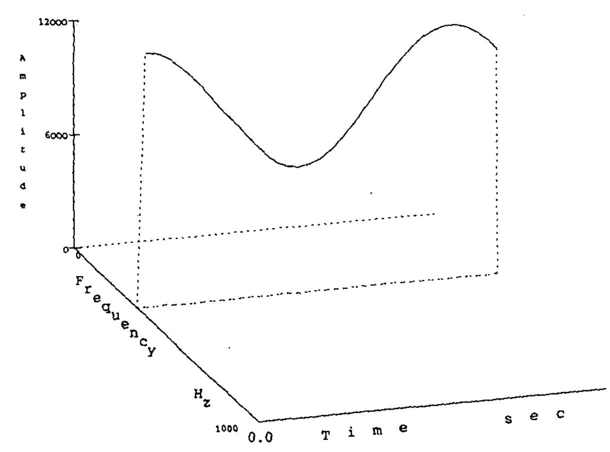


Fig. 4. Nearest-neighbor extrapolation with linear segment to fill gap in Fig. 3.

the MQ analysis of Fig. 7.

Several different 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 arbitrary—but often reasonable—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. If the tracks cannot be matched satisfactorily using the prediction, the track is cross-faded with the matched track to avoid any phase discontinuity.

Step 3: Using the track-matching information from step 2, cubic polynomial extrapolation functions are generated for amplitude and phase across the gap. The cubic amplitude function is obtained simply by solving for the coefficients of the cubic function that passes through the two points preceding and the two points following the gap, while the cubic phase function is obtained using the measured phase and frequency (phase derivative) for the frame preceding and the frame following the gap.

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 Fig. 7 is depicted in Fig. 8. The synthesized time-domain signal in the gap interval is shown in Fig. 9.

Another example of the extrapolation procedure is presented in Figs. 10 and 11. The signal in this case is a segment of a Mozart orchestral work, demonstrating the behavior of the procedure on complex polyphonic material.

4 DISCUSSION

Although the MQ procedure provides an elegant representation for the gap extrapolation problem, there are several limitations built into the assumptions of both the sinusoidal model and the gap-filling strategy. The limitations are related primarily to the matching that

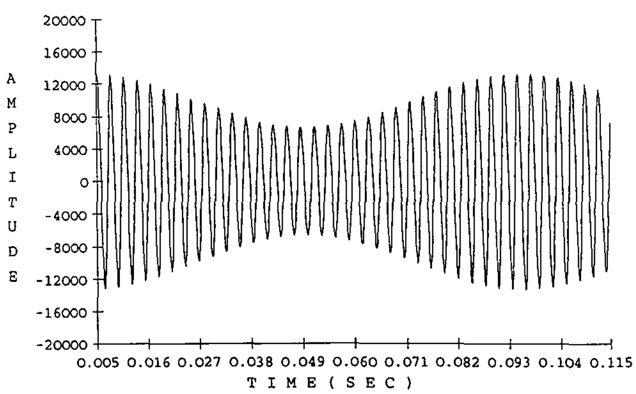


Fig. 5. Reconstructed signal with gap fill resynthesized from extrapolated MQ analysis data of Fig. 4.

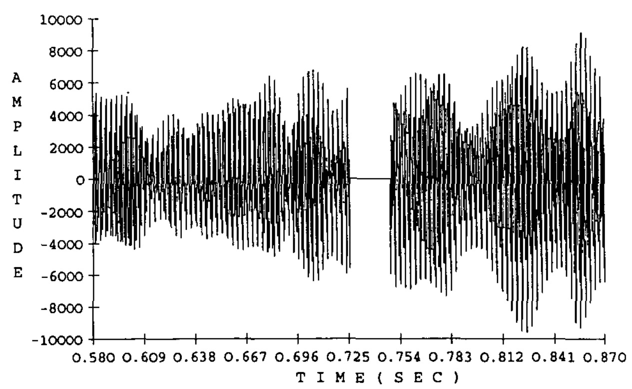


Fig. 6. Example signal with 22.7-ms artificial gap (1000 samples at 44.1-kHz sample rate). Signal is from recording of a soprano singer.

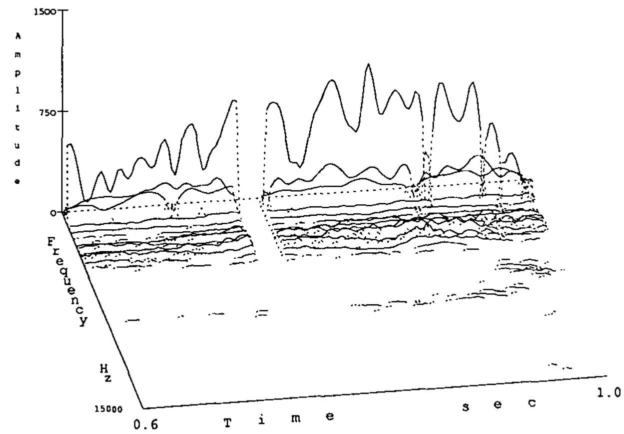


Fig. 7. MQ analysis of signal of Fig. 6, with gap.

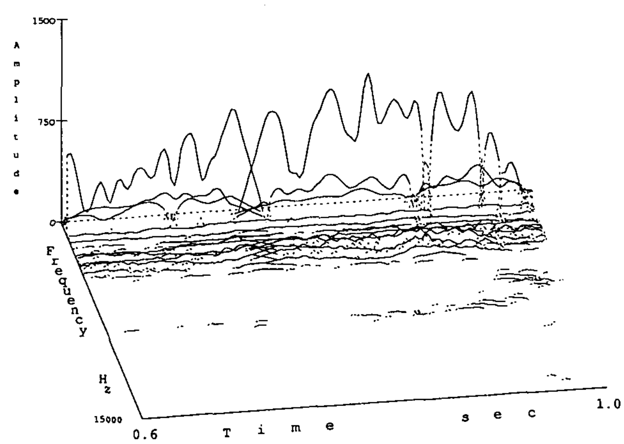


Fig. 8. Cubic function polynomial extrapolation to fill gap in Fig. 7.

must be done between the track information preceding and following the gap: if the spectrum is changing rapidly within the gap interval, the segmental extrapolation may be unsatisfactory.

4.1 Continuity of Frequency Tracks

One important consideration for generating a seamless gap extrapolation is to maintain amplitude and phase continuity of each frequency track. First-order track amplitude continuity can be guaranteed by the cubic function approach, but phase continuity can be problematic. For example, consider the simplified situation depicted in Fig. 12. In this example the MQ analysis of the signal both before and after the gap consists of just two frequency tracks, whereas dozens of tracks are typically present in real analysis data. If the simple nearest-neighbor frequency track matching algorithm is used, the upper track before the gap is matched to the lower track after the gap, while the prior lower track "dies" and the subsequent upper track is "born." If, on the other hand, the tracks are matched to minimize the rate of frequency change, the lower and upper tracks before the gap are matched to the lower and upper tracks after the gap. The choice of matching method affects the phase relationships among the partials because the frequency and phase are related by integration. Thus the degree to which the synthesized signal in the gap matches the signal at the end of the gap depends on how the track matching takes place. For short gaps (100–200 samples) the discontinuity at the end of the gap is often negligible. For longer gaps a simple cross-fade from the synthesized signal to the actual signal following the gap works well in practice. An alternative procedure to perform the track matching using an iterative discontinuity minimization constraint is currently being investigated.

4.2 Threshold Range and Spectral Tilt

A subtle but important consideration for Fourier-transform-based analysis techniques is the degree to which low-level spectral components can be identified in the presence of stronger components at other frequencies. The range of resolvable component levels can be referred to as the *threshold range* of the analysis procedure. The ability to identify low-level components is related to the window function used in the discrete Fourier transform, and typifies the well-known tradeoff between the desire to reduce main-lobe width in order to identify closely

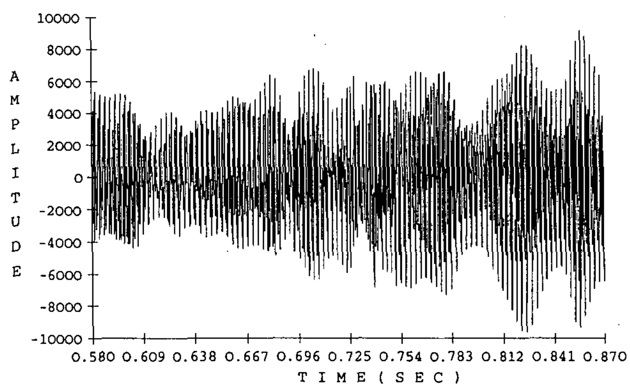


Fig. 9. Reconstructed signal with gap fill resynthesized from extrapolated MQ analysis data of Fig. 8.

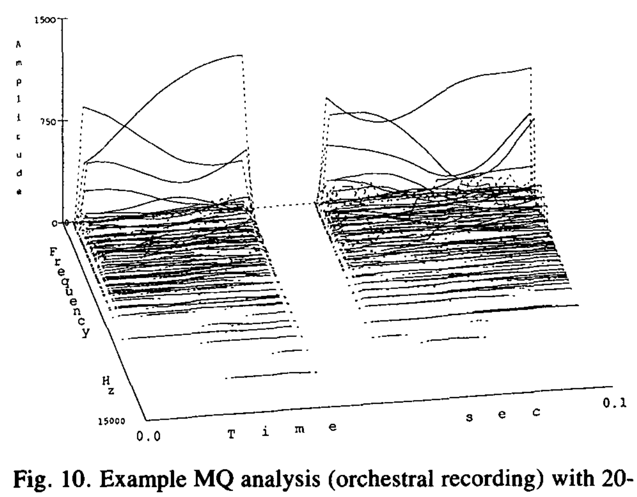


Fig. 10. Example MQ analysis (orchestral recording) with 20-ms gap.

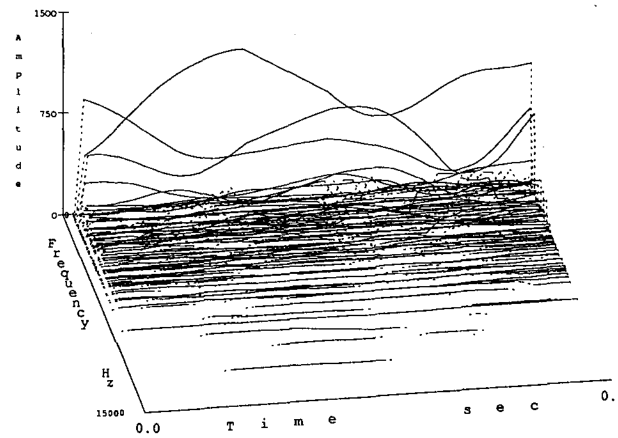


Fig. 11. Reconstructed MQ analysis with gap fill for analysis example of Fig. 10.

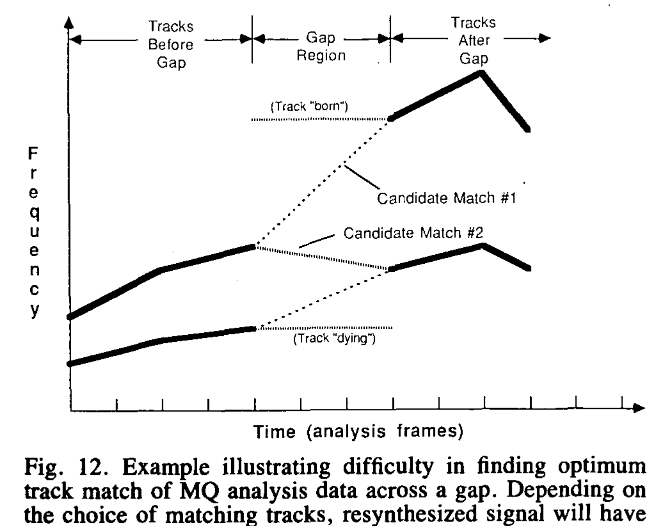


Fig. 12. Example illustrating difficulty in finding optimum track match of MQ analysis data across a gap. Depending on the choice of matching tracks, resynthesized signal will have a different time-domain shape due to integral relationship between component phase and frequency.

spaced components with similar amplitude, and the desire to reduce sidelobe level to identify widely spaced components with very different amplitudes. With common analysis windows the difference between the mainlobe and the highest sidelobe levels is between 30 and 60 dB [21].

An elementary example comparing the main-lobe width and the sidelobe behavior of two different window functions is shown in Fig. 13. The Hamming window transform (solid line) has a more narrow main lobe than the Blackman window (dotted line), but the Blackman sidelobes are lower in level. This tradeoff can be made manually according to the needs of the particular situation.

The analysis threshold range is important in the proposed extrapolation procedure because of the characteristic low-pass *spectral tilt* exhibited by many common musical signals. In other words, the high-frequency portion of the spectrum generally contains components with significantly lower amplitude than the low-frequency portion. The result is that the analysis sidelobes due to strong low-frequency components can obscure the presence of the high-frequency spectral peaks. Therefore in practice it is necessary to set a minimum threshold in the MQ analysis to prevent the misidentification of sidelobes as actual signal components in the spectral magnitude.

If the threshold is set high enough to avoid the analysis sidelobes, the MQ procedure will contain only the very strongest high-frequency components. Unfortunately this can result in an audible deficit of high-frequency material during the reconstructed gap interval. It is possible, however, to counteract the low-pass character of musical signals to some extent by applying a high-pass preemphasis to help flatten the spectrum of the signal prior to the MQ analysis, followed by a complementary deemphasis of the synthesized output signal. In any case, operator intervention is typically needed to coordinate the analysis procedure.

4.3 Performance for Noiselike Signal Components

As mentioned previously, the MQ analysis procedure models the input signal as the sum of many independent sinusoidal components. If the input signal contains broad-band noise or noiselike components, such as fricative speech sounds, percussion, or bow scrape, the noise is represented in the MQ analysis as a collection of sinusoidal components with rapidly varying amplitude and frequency, as shown in Fig. 14. Although the quality of a noisy signal synthesized by the MQ procedure is usually quite acceptable, the situation for reconstructing gaps containing noisy components is not addressed by the simple component matching procedures described in Section 3. This is because the explicit assumption of a slowly time-varying quasi-harmonic spectral envelope is violated in the case of broad-band noise.

Several possibilities for extending the MQ gap extrapolation strategy to noisy signals are currently under consideration, including the deterministic/stochastic separation of Serra and Smith [15], [16]. Such a scheme would require the characteristics of the noisy material to be estimated separately from the more slowly varying sinusoidal components.

5 CONCLUSION

In this engineering report a method for estimating missing or corrupted samples in a digital audio data stream was presented. 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 engineering report is very effective for concealing gaps of up to 30 ms in duration, although much longer gaps can be concealed if the signal spectrum remains relatively constant during the gap in-

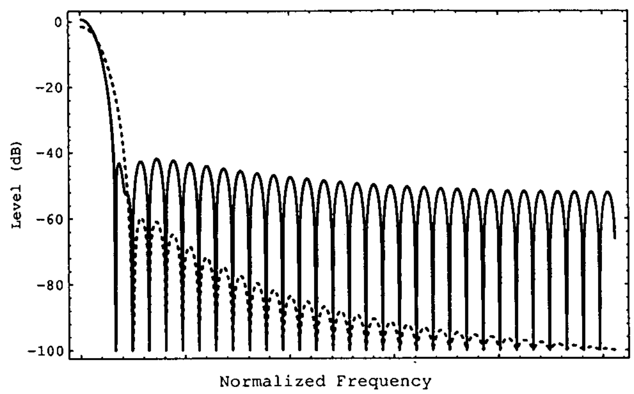


Fig. 13. Window function sidelobe behavior. Solid line (upper)—Hamming window; dotted line (lower)—Blackman window.

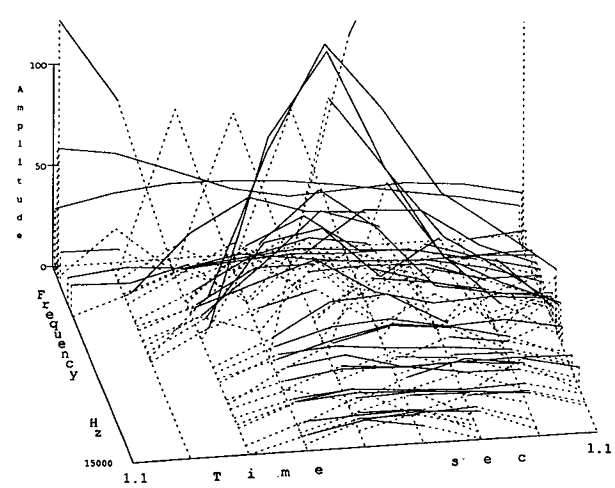


Fig. 14. MQ analysis of noisy signal.

terval. Additional work is needed to address the problem of noisy signal components within the sinusoidal analysis framework.

6 ACKNOWLEDGMENT

This work was supported in part by the Research Council and the Department of Electrical Engineering, University of Nebraska–Lincoln. The author would also like to acknowledge the helpful comments and suggestions of the review board.

7 REFERENCES

[1] S. V. Vaseghi and R. Frayling-Cork, "Restoration of Old Gramophone Recordings," *J. Audio Eng. Soc.*, vol. 40, pp. 791–801 (1992 Oct.).

[2] R. W. Schafer and L. R. Rabiner, "A Digital Signal Processing Approach to Interpolation," *Proc. IEEE*, vol. 61, pp. 692–702 (1973).

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

[4] A. Papoulis, "A New Algorithm in Spectral Analysis and Bandlimited Extrapolation," *IEEE Trans. Circuits Sys.*, vol. CAS-22, pp. 735–742 (1975).

[5] J. A. Cadzow, "An Extrapolation Procedure for Bandlimited Signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-27, pp. 4–12 (1979)

[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] 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).

[8] A. J. M. Janssen, R. N. J. Veldhuis, and L. B. Vries, "Adaptive Interpolation of Discrete-Time Signals that Can Be Modeled as Autoregressive Processes," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, pp. 317–330 (1986).

[9] E. J. Diethorn and D. C. Munson, "A Linear, Time-Varying System Framework for Noniterative Discrete-Time Bandlimited Signal Extrapolation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-39, pp. 55–68 (1991)

[10] J. S. Abel and J. O. Smith, "Restoring a Clipped Signal," in *Proc. 1991 IEEE ICASSP* (1991), pp. 1745–1748, paper D4a.7.

[11] R. J. McAulay and T. F. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, pp. 744–754 (1986).

[12] J. O. Smith and X. Serra, "PARSHL: An Analysis/Synthesis Program for Nonharmonic Sounds Based on a Sinusoidal Representation," in *Proc. 1987 Int. Computer Music Conf.* (Computer Music Assn., San Francisco, CA, 1987), pp. 290–297.

[13] R. C. Maher and J. W. Beauchamp, "An Investigation of Vocal Vibrato for Synthesis," *Appl. Acoust.*, vol. 30, pp. 219–245 (1990).

[14] R. C. Maher, "Evaluation of a Method for Separating Digitized Duet Signals," *J. Audio Eng. Soc.*, vol. 38, pp. 956–979 (1990 Dec.).

[15] X. Serra, "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).

[16] X. Serra and J. O. Smith, "Spectral Modeling Synthesis: A Sound Analysis/Synthesis System Based on a Deterministic plus Stochastic Decomposition," *Computer Music J.*, vol. 14, pp. 12–24 (1990 Winter).

[17] T. F. Quatieri and R. J. McAulay, "Speech Transformations Based on a Sinusoidal Representation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, pp. 1449–1464 (1986).

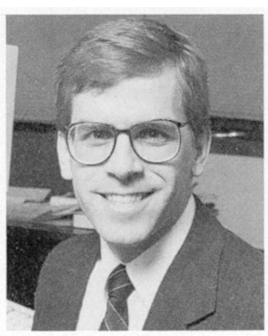
[18] R. C. Maher, "Sinewave Additive Synthesis Revisited," presented at the 91st Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 39, p. 996 (1991 Dec.), preprint 3128.

[19] M. R. Portnoff, "Implementation of the Digital Phase Vocoder Using the Fast Fourier Transform" *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-24, pp. 243–248 (1976).

[20] M. Kahrs, private communication (1993).

[21] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1989).

THE AUTHOR



Robert C. (Rob) Maher was born in Cambridge, UK, in 1962. He holds a B.S. degree from Washington University, St. Louis, MO (1984), an M.S. degree from the University of Wisconsin–Madison (1985), and a Ph.D. degree from the University of Illinois-Urbana (1989), all in electrical engineering. While a student, he worked as a graduate research assistant for the University of Illinois Computer Music Project. He is currently an assistant professor of electrical engineering at the University of Nebraska–Lincoln, with teaching and research interests in the application of advanced digital signal processing methods in the fields of audio engineering, electroacoustics, and computer music. He also performs limited consulting work in the digital audio signal processing field.

Dr. Maher is a member of several professional organizations, including the Audio Engineering Society, Institute of Electrical and Electronics Engineers, ASA, ASEE, ICMA, and the Tau Beta Pi, Eta Kappa Nu, and Phi Kappa Phi honor societies.