SHORT-TIME HOMOMORPHIC ANALYSIS*

José M. Tribolet, Thomas F. Quatieri and Alan V. Oppenheim

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, MA 02139

Abstract

Homomorphic deconvolution has been successfully applied in a variety of areas. In many cases of interest, including speech and seismic processing, the signals to be analyzed are non-stationary and approximately follow a convolutional model only on a short-time basis. Thus, a window is applied to the data.

In this paper a first attempt is made to understand the interaction between short-time windowing and homomorphic deconvolution. A model for short-time homomorphic analysis is proposed, which provides a framework for the interpretation of window effects encountered in speech and seismic data processing.

I. INTRODUCTION

Homomorphic deconvolution has been successfully applied in a variety of areas [1]. It relies on the property that the complex cepstrum maps convolution to addition and that in many applications, the complex cepstra of the components to be deconvolved occupy approximately disjoint time (or quefrency) regions.

In many cases of interest, including speech and seismic processing, the signals to be analyzed are non-stationary and approximately follow a convolutional model only on a short-time basis. Thus, a window is applied to the data. The classes of windows conventionally used for homomorphic deconvolution have been Hamming (and similar) windows for speech processing, and exponential windows for seismic processing. Although the use of these windows has led in many instances to good performance, as in the cases of minimum phase speech analysis and seismic dereverberation in shallow water, recent attempts to use homomorphic signal processing in other contexts have demonstrated the importance of the short-time window in the potential determination of the signal components by homomorphic deconvolution.

In this paper a first attempt is made to understand the interaction between short-time windowing and homomorphic deconvolution. Toward this end in Section II, a model for the short-time homomorphic analysis is proposed. While this representation is not unique it appears useful in understanding window effects as well as providing a framework within which the cepstral separation of the signal components might be tested and even improved. In Section III we shall discuss some of the effects of windows in speech and seismic data processing.

II. A SHORT-TIME MODEL FOR HOMOMORPHIC DECONVOLUTION

Consider a segment of data $x(n)$ which on a short-time basis satisfies a convolutional model and to which a window has been applied so that $x(n)$ can be expressed in the form

$$x(n) = \{p(n) * r(n)\}\, w(n) \qquad (1)$$

where $p(n)$ and $r(n)$ are the two convolutional components and $w(n)$ is the window. In carrying out homomorphic deconvolution, the complex cepstrum is assumed to consist of the sum of two components, occupying disjoint time intervals, and the basis for deconvolution is time gating of the complex cepstrum to recover the desired component. This corresponds to assuming that $x(n)$ consists of a convolution of components which is, of course, not strictly true for the windowed segment represented by eq. (1). Furthermore, there is the underlying assumption that the low time and high time portions of the complex cepstrum can be associated with $p(n)$ and $r(n)$ in eq. (1). Thus, the window is assumed slowly varying compared with $p(n)$ so that $x(n)$ is approximated as

$$x(n) \simeq p(n) * [r(n) \cdot w(n)] \qquad (2)$$

This implies among other things, that the estimate of $p(n)$ obtained by homomorphic deconvolution will be relatively insensitive to the shape and time registration of the window. While for minimum phase speech analysis using the real cepstrum this is an acceptable assumption [2], it is generally

*This work was supported in part by the Advanced Research Projects Agency monitored by ONR under Contract N00014-75-C-0951-NR 049-308 and in part by the National Science Foundation under Grant ENG71-02319-A02.

not for mixed-phase analysis using the complex cepstrum. Consider for example the artificial speech waveform x[n] in Fig. 1c, derived from the convolution of a pulse p(n) and an impulse train r(n) in Figs. 1a and 1b, respectively. The pulse

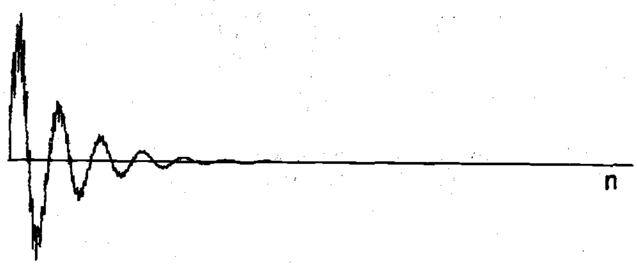

Figure 1a: Artificial Vocal Tract Response

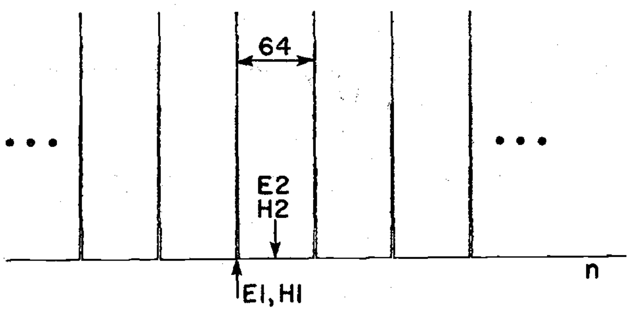

Figure 1b: Excitation Function

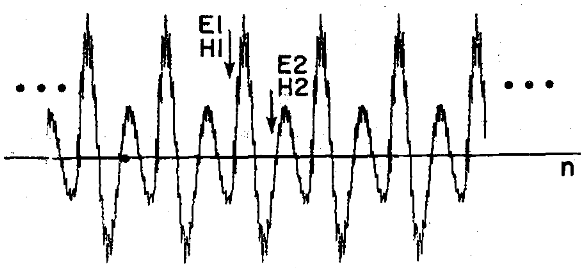

Figure 1c: Artificial Speech Waveform

p(n), representing a synthetic vocal tract response is mixed phase with two minimum phase pole pairs and one zero pair, which is outside the unit circle. The impulse train r(n) is a synthetic vocal cord excitation with a 64-sample period. Let us consider the result of homomorphic deconvolution by low-time and high-time gating of the complex cepstrum with different windows and onset time. In terms of the model of eq. (2), low-time gating should recover the basic pulse p(n) and high-time gating the windowed impulse train.

Figures 2a, b and c demonstrate the effect of an exponential window of 400 point duration and decay rate of .992. Its onset is indicated in Figs. 1b and 1c by the arrow E1. The low-time cepstral window cut off T_c equals 46 and the correspondent low-time and high-time estimates, $x_L(n)$ and $x_H(n)$, in Figs. 2b and 2c can be seen to be very close to the pulse p(n) and the windowed impulse train r'(n)=w(n)r(n) as predicted by (2).

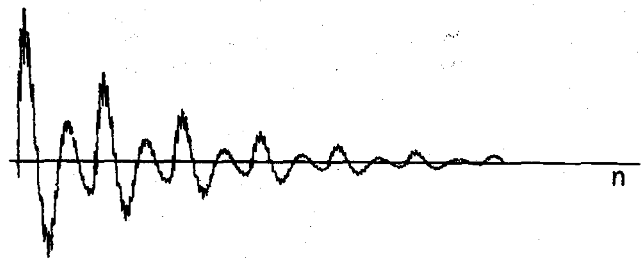

Figure 2a: Speech waveform with an exponential window of onset E1

Figure 2b: $x_L(n)$ derived from 2a

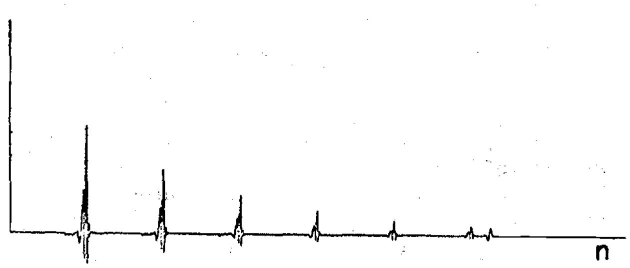

Figure 2c: $x_H(n)$ derived from 2a

In Figures 3a, b, and c the effect of window onset is illustrated. The exponential window used in Fig. 2 has been shifted 32 points as depicted in Fig. 3a. The new window onset is indicated in Figs. 1b and 1c by the arrow E2. Now, although it is still reasonable to associate $x_H(n)$ in Fig. 3c with r'(n), it is no longer possible to associate the estimated low-time pulse $x_L(n)$ in Fig. 3b with the original pulse p(n) of Fig. 1b.

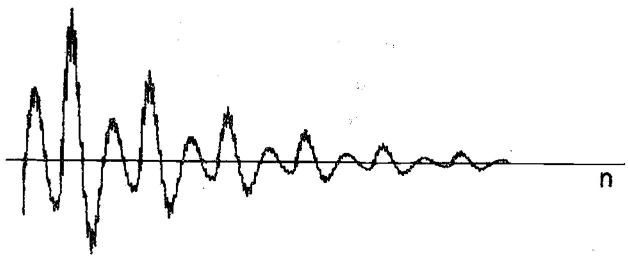

Figure 3a: Speech waveform with an exponential window of onset E2

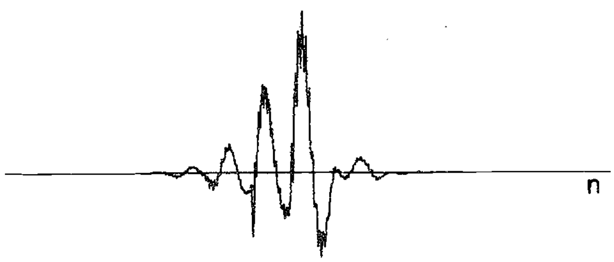

Figure 3b: $x_L(n)$ derived from 3a

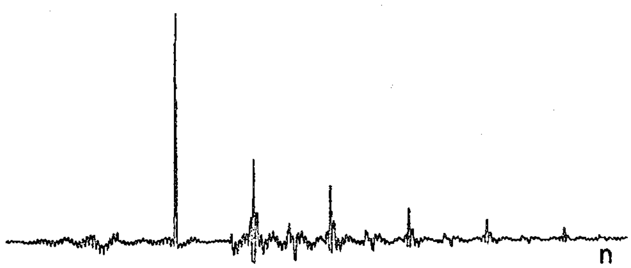

Figure 3c: $x_H(n)$ derived from 3a

Next, the effects of a conventional Hamming window (of length 256 and onset indicated by arrow H2 in Figs. 1b and 1c) are demonstrated in Figs. 4a, b, and c. Once again $x_H(n)$ in Fig. 4c matches $r(n)$, but $x_L(n)$ differs significantly from $p(n)$ and from the previously estimated pulse of Fig. 3b.

A similar effect is exhibited with the convolution of a basic pulse with a non-periodic impulse train as is the case for example with seismic data. Consider the artificial seismic trace $x(n)$ of Fig. 5c, derived from the convolution of the airgun signature of Fig. 5a with the synthetic earth impulse response of Fig. 5b. The airgun signature is a mixed phase pulse and the earth impulse response is a non-periodic impulse train with random amplitudes and arrival times. Once again, let us perform homomorphic deconvolution on different intervals of $x(n)$ using different windows.

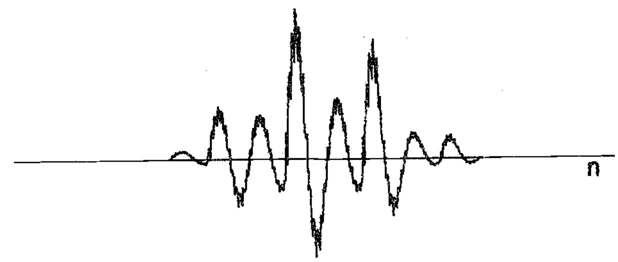

Figure 4a: Speech waveform with a Hamming window of onset H2

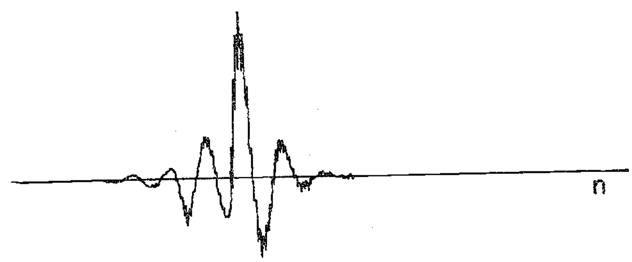

Figure 4b: $x_L(n)$ derived from 4a

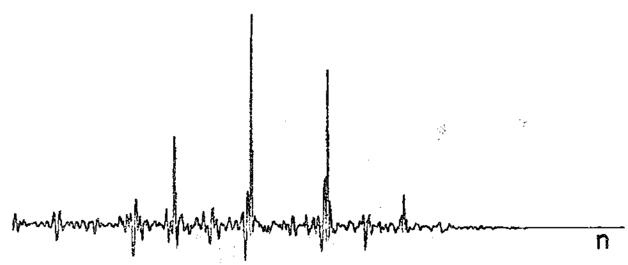

Figure 4c: $x_H(n)$ derived from 4a

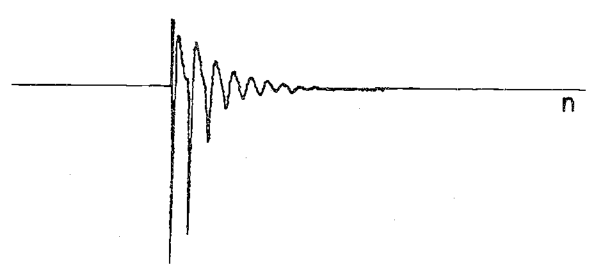

Figure 5a: Seismic airgun signature

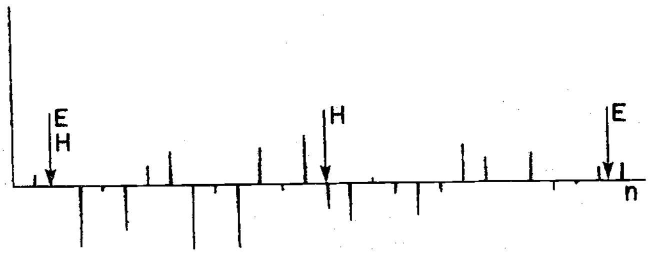

Figure 5b: Artificial earth impulse response

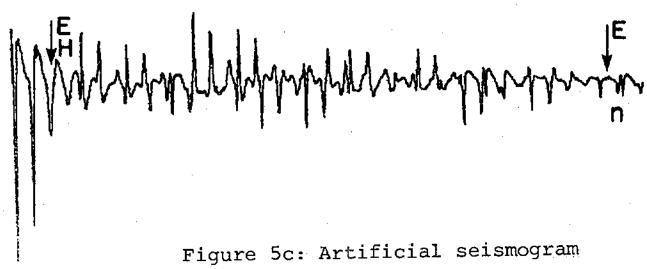

Figure 5c: Artificial seismogram

In Figs. 6a, b, and c the use of an exponential window, with decay .994 is demonstrated; the window onset and offset are shown in Fig. 5 by the arrows E. The resultant segment is shown in Fig. 6a. Fig. 6b shows the $x_L(n)$ obtained by low-time gating and 6c the estimate $x_H(n)$ obtained by high-time gating. Once more, $x_H(n)$ closely matches $r(n) \cdot w(n)$, as seen in Fig. 6d, after exponential deweighting of $r'(n)$. However, the estimated pulse $x_L(n)$ is clearly very different from the original signature $p(n)$.

Finally, consider Figs. 7a, b, and c. In Fig. 7a, a segment obtained by Hamming windowing $x(n)$, in Fig. 5a, with the window onset and offset indicated by the H arrows. The high-time estimate $x_H(n)$ in Fig. 7c, is seen to match $r'(n)$, but the low-time estimate $x_L(n)$ in Fig. 7b is once again different from the airgun pulse, and also from the pulse of Fig. 6b.

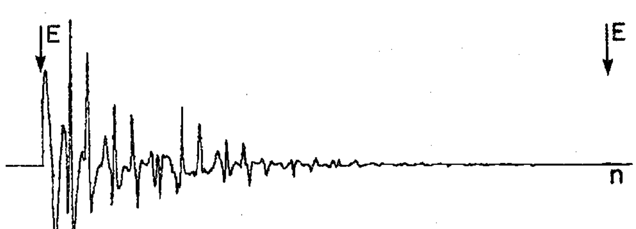

Figure 6a: Seismogram segment with an exponential window with onset and offset E

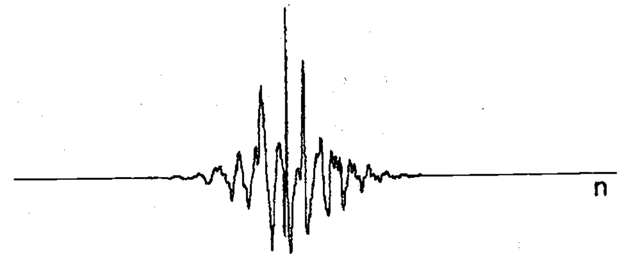

Figure 6b: $x_L(n)$ derived from 6a

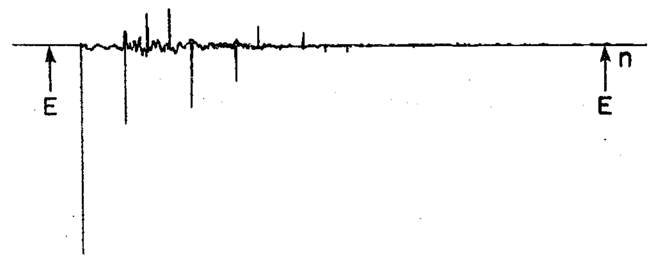

Figure 6c: $x_H(n)$ derived from 6a

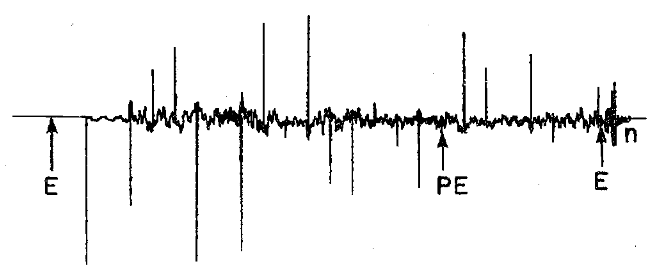

Figure 6d: $x_H(n)$ after exponential deweighting

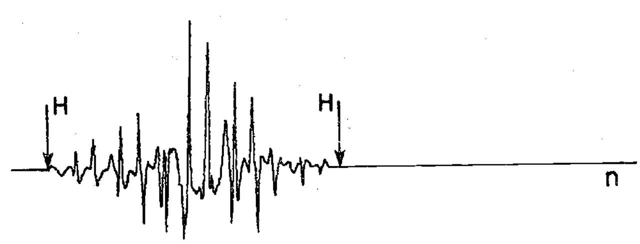

Figure 7a: Seismogram segment with a Hamming window with onset and offset H

Figure 7b: $x_L(n)$ derived from 7a

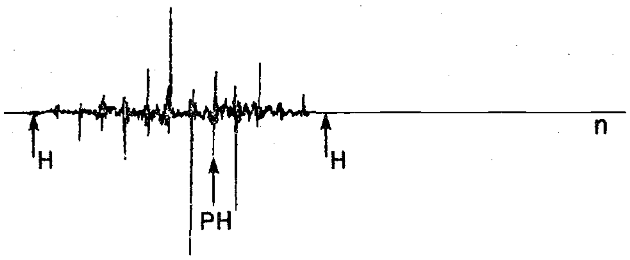

Figure 7c: $x_H(n)$ derived from 7a

The principle effect that we observe in both sets of examples is the sensitivity of the estimated pulse to the time registration and the shape of the window. In retrospect this is not surprising since there are many different pulse shapes which, when convolved with an impulse train will result in the same composite signal. This suggests modifying eq. (2) to

$$x(n) = p'(n) * r'(n) = p'(n) * [r(n)w(n)] \qquad (3)$$

where $p'(n)$ is not necessarily the pulse used to generate the data and is no longer independent of the data structures, window shape, onset and duration. If in carrying out homomorphic deconvolution an accurate estimate of the original pulse shape is required then the window shape and time registration must be chosen carefully. For some applications, however, the true pulse shape is not important. In a speech analysis-synthesis system, for example, the pulse obtained in the analysis is convolved with an impulse train in the synthesis, and it is the quality of the reconstructed speech that is important. In seismic processing, it is often the reflector series that is of interest, and when the pulse $p(n)$ is estimated it is used to design an inverse filter to obtain the reflector series. In these cases the short-time model of eq. (3) may be acceptable even when $p'(n)$ is very different from $p(n)$.

In the next section, we discuss these two classes of examples in more detail focussing in particular on the effects of windowing on the estimation of the signal components and the possible implications.

III. SHORT-TIME HOMOMORPHIC ANALYSIS OF SPEECH AND SEISMIC DATA

Seismic Processing

A common goal in seismic data processing is the recovery of the impulse train, representing the earth reflector series. In most cases, however, little is known about the data except that the reflector series $r(n)$ is, ideally, composed of a series of sharp impulses with random arrival times and amplitudes, and that the seismic wavelet $p(n)$ is known to be smooth and short, compared with $r(n)$.

Since, in general, $r'(n)$ is mixed-phase, its complex cepstrum $\hat{r}'(n)$ will have contributions for both positive and negative values of n. This may cause significant low-time interference between their complex cepstra $\hat{r}'(n)$ and $\hat{p}'(n)$ which must be accounted for in subsequent analysis.

The approach taken in seismic processing has been to use an exponentially decaying window $w(n)$ such that $r'(n)$ becomes minimum phase. This eliminates the low-time cepstral interference between the signal components up to the cepstral time $T_c = T_\Delta$, where T_Δ is the inter-arrival time between the first two impulses of $r(n)$.

However, it often happens that such T_Δ is so low, that the high-time interference is still significant. As a result, earth impulse response estimates based on the high-time cepstral estimate $x_H(n)$ often display a significant convolutional noise component which may seriously degrade the estimate.

We have purposely attempted to use a variety of windows for short-time seismic deconvolution: exponential decreasing and increasing, Hamming, rectangular, Rayleigh, etc. Once a window, yielding a reasonable estimate of $p'(n)$ has been found, we design from such an estimate a linear time-invariant inverse filter which is ultimately used to process the seismic segment $x(n)$.

A number of filter design techniques have been used, namely Homomorphic Prediction [3] and Wiener Filtering [4]. We illustrate the use of Wiener Filtering in Fig. 8. In Fig. 8a the output response of an Optimum Lag Wiener Spiking Filter, of length 50, designed from the estimate of $p'(n)$ in Fig. 6b is shown; the input to the filter is the seismic segment of Fig. 6a, without the exponential weighting. This deconvolved seismogram is of high resolution and is only slightly better than the high-time cepstral estimate of Fig. 6d (compare resolution in the interval indicated by the arrow PE in both Figs. 6d and 8a).

Consider next the output response of an Optimum Lag Wiener Spiking Filter, also of length 50, designed from the estimate of $p'(n)$ in Fig. 7b. The output responses shown in Figs. 8b and 8c correspond, respectively, to inputs which are the seismic segment of Fig. 8a, with and without the

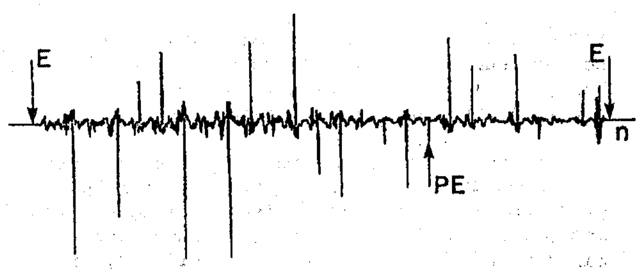

Figure 8a: r(n) estimate for segment of Fig. 6a

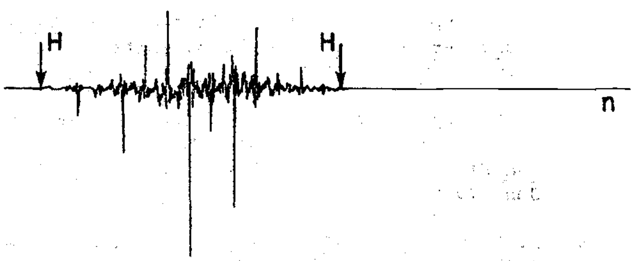

Figure 8b: r'(n) estimate for segment of Fig. 7a

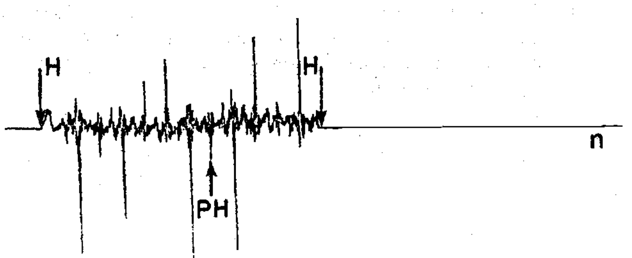

Figure 8c: r(n) estimate for segment of Fig. 7a

Hamming window. Note that the output in Fig. 8b can be quite accurately derived from Fig. 8c by Hamming windowing, which is consistent with the model of eq. (3). Here, the improvement over the high-time estimate of Fig. 7c is more noticeable. For example, the reflection indicated by the arrow PH is accurately estimated in Fig. 8c whereas in Fig. 7c it appears as a doublet.

Our experience suggests that the use of inverse filters in conjunction with homomorphic wavelet estimation has enormous potential for seismic deconvolution. In particular, the use of Wiener filters, designed from the appropriate low-time cepstral estimates of p'(n), have performed rather well and seem to overcome the bothersome convolutional noise phenomenon so often associated with the high-time cepstral estimates of r'(n).

Speech Processing

A goal in speech data processing is the recovery of the vocal tract impulse response p(n) or, if we restrict our attention to vocoder applications, the recovery of the pulse p'(n) from which good quality speech synthesis might be attained.

Let us first discuss the analysis-synthesis problem and, in particular, the consideration of voiced speech. Here, the cepstrum of r'(n) is always high time due to its pitch periodicity. Thus, the low-time cepstral estimate $x_L(n)$ will usually accurately estimate p'(n). From eq. 6 it follows then that the synthesis of the short-time segment x(n) can be accomplished, except for its exact relative frame positioning in time. Thus, to avoid pitch jitter, synchronization between speech frames seems to be required. This problem, at the present time, is under an active investigation where we are seeking a simple method of properly concatenating adjacent speech segments.

Estimation of the true vocal tract impulse response p(n) is an extremely difficult task for it requires precise estimation of logmagnitude and phase characteristics. From examination of both analytical and computer generated examples it appears that the logmagnitude can be strongly dependent on the short time window. For example, Fig. 9a shows the superposition of the logmagnitude curves associated with Figs. 1c, 2c and 3c. Here we see that the pulse synchronized window onset yields the most accurate approximation to the true logmagnitude. Also, in Fig. 9b we see a comparison of the results obtained from the 256 point Hamming window. Included here is the logmagnitude due to a Hamming window with onset indicated by arrow H1 in Figs. 1b and 1c.

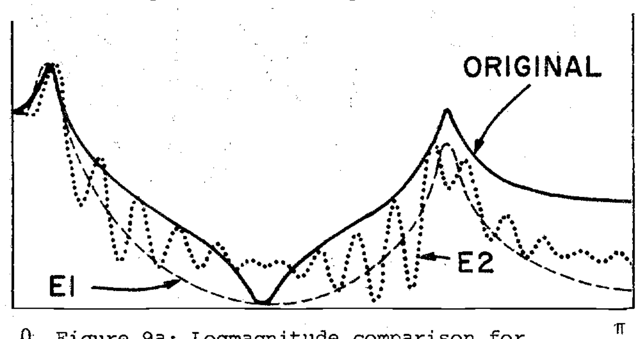

Figure 9a: Logmagnitude comparison for exponential windows

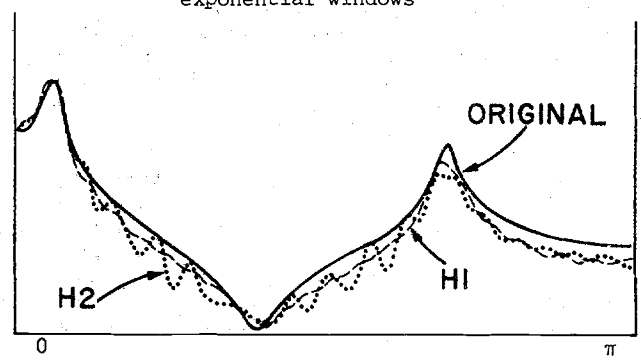

Figure 9b: Logmagnitude comparison for Hamming windows

Figures 10a and 10b show similar comparisons of the phase curves obtained from the various windowed intervals. Again, we see quite significant differences among the estimates for the different onsets and window shapes.

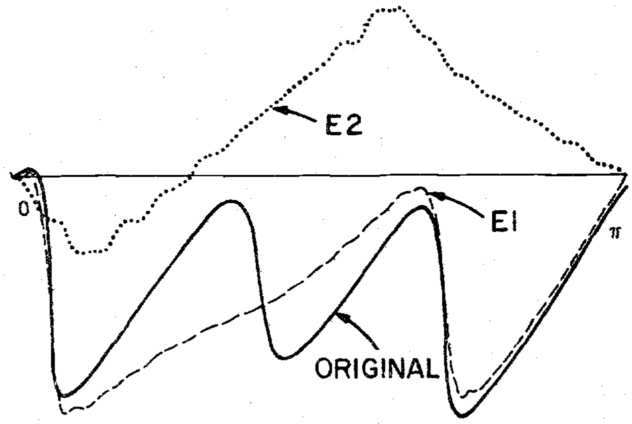

Figure 10a: Phase comparison for exponential windows

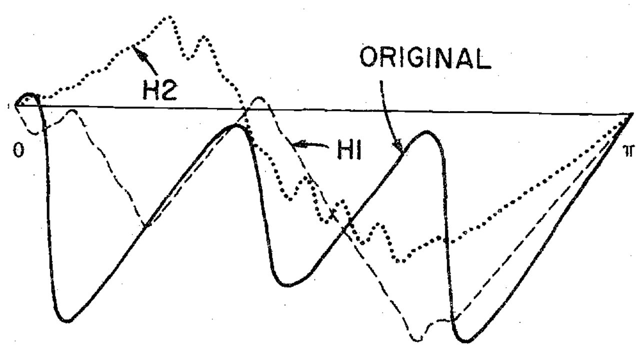

Figure 10b: Phase comparison for Hamming windows

Although a complete understanding of these effects is lacking, some heuristic insights have been gained. In general, it appears that the window w(n) affects the convolutional model such that the logmagnitude and the phase characteristics of p'(n) can differ significantly from those of the original pulse.

Analysis of mathematically tractable signals and windows have led us to speculate on two possible causes for this. First, windowing may smooth the logmagnitude and phase curves near high Q poles and zeros. Although this effect is localized for the logmagnitude characteristic, the phase curve is globally affected and it can drastically change shape. For example, Figs. 9a and 10a demonstrate this for characteristics due to the synchronized exponential window (arrow E1). The logmagnitude exhibits a smoothing of the zero of the original pulse, while approximately maintaining the correct shape. The phase curve, on the other hand, has little resemblance to the original curve in the region of the zero.

Secondly, window shape and onset may introduce additional zeros whose topology can be detrimental for accurate pulse estimation. Some analytical results, for example, have yielded minimum and maximum phase zeros uniformly spaced close to the unit circle. This is a possible explanation for the ripples in the logmagnitude of Figs. 9a and 9b.

References

[1] A. V. Oppenheim and R. W. Schafer, Digital Signal Processing, Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1970, Chapter 10

[2] A. V. Oppenheim and R. W. Schafer, "Homomorphic Analysis of Speech," IEEE Trans. Audio and Electroacoust., vol. AU-16, pp. 221-226, June 1968

[3] A. V. Oppenheim, G. E. Kopec and J. M. Tribolet, "Signal Analysis by Homomorphic Prediction," IEEE Trans. on Acoustics, Speech and Signal Processing, vol. ASSP-24, no. 4, pp. 327-332, August 1976

[4] S. Treitel and E. A. Robinson, "The Design of High-Resolution Digital Filters," IEEE Trans. on Geoscience Electronics, vol. GE-4, no. 1, pp. 25-38, June 1966