GENERALIZED TRANSFER FUNCTION ESTIMATION USING EVOLUTIONARY SPECTRAL DEBLURRING

S. I. Shah, L. F. Chaparro and A. El-Jaroudi

348 Benedum Hall
Department of Electrical Engineering
University of Pittsburgh,
Pittsburgh, PA 15261

ABSTRACT

We present a new technique for estimating the generalized transfer function (GTF) of a time–varying filter from time–frequency representations (TFRs) of its output. We use the fact that many of these representations can be written as blurred versions of the GTF. The approach consists in minimizing the error between the TFR found from the data and that found by blurring the GTF. The problem as such has many solutions. We, therefore, additionally constrain it to minimize the distance between the GTF–based spectrum and the autoterms of the Wigner distribution, suppressing the cross terms using an appropriate signal dependent mask function. To illustrate the performance of the proposed procedure we apply it to the spectral representation of speech and to signal masking and demonstrate its superior performance over the existing methods.

1. INTRODUCTION

There is an increasing interest in time–frequency analysis techniques for the characterization of the time-dependent spectra of signals found in practical applications [1]. Time–frequency analysis is related to the theory of linear time-varying systems (LTV) by extending the spectral representation of stationary signals to the non–stationary case. According to the Wold-Cramer representation [8, 9], a non-stationary signal $x(n)$ can be considered the output of a causal LTV system,

$$x(n) = \int_{-\pi}^{\pi} H(n,\omega) e^{j\omega n} dZ(\omega) \qquad (1)$$

where $H(n,\omega)$ is the generalized transfer function (GTF) [2] of the LTV system evaluated on the unit circle and $Z(\omega)$ is an orthogonal increments process. The Wold-Cramer evolutionary spectrum $S_{ES}(n,\omega)$ of $x(n)$ is then obtained by considering the time-dependent variance [8, 9] of $x(n)$, yielding

$$S_{ES}(n,\omega) = |H(n,\omega)|^2. \qquad (2)$$

Detka et al. [10] have shown that the spectrogram [2] and Cohen's class of bilinear distributions [4] are related to the evolutionary spectrum and the generalized transfer function. In fact, by substituting the expression in (1) into the general form of the bilinear distributions given by

$$S_{BD}(n,\omega) = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} W[l-n,k] \cdot$$
$$x^*(l-k,\lambda) x(l+k,\lambda) e^{-j(\omega-\lambda)2k} d\lambda, \qquad (3)$$

where $W[.,.]$ is a weight function with finite support, we find that these distributions are related to the generalized transfer function by

$$E\{S_{BD}(n,\omega)\} = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} W[l-n,k] \int_{-\pi}^{\pi} H^*(l-k,\lambda) H(l+k,\lambda) e^{-j(\omega-\lambda)2k} d\lambda \qquad (4)$$
$$= \sum_{k=-\infty}^{\infty} \int_{-\pi}^{\pi} S_{ES}(k,\lambda) G(n-k,\omega-\lambda;H) \quad d\lambda \qquad (5)$$

where $E\{.\}$ is the expected value operator and $G(\cdot,\cdot;\cdot)$ is a blurring function that depends on $W[\cdot,\cdot]$ and $H(\cdot,\cdot)$. In this paper, we exploit this relationship to compute improved estimates of the GTF and the ES using a deconvolution approach. The deconvolution problem of interest consists in, given a signal $x(n)$ and a bilinear distribution, we wish to obtain a $H(n,\omega)$, such that when blurred according to (4) it results in the given bilinear distribution.

Pitton et al. [6], who proposed a similar approach using the spectrogram, circumvent the dependence of the blurring function on the GTF by assuming that the process under consideration is stationary within the spectrogram window. While such an assumption simplifies the blurring functions, it can result in an incomplete deblurring. In our method, we do not assume stationarity and differ also with the current deconvolution techniques [5, 6] in the sense that: (a) our technique can be based on any TFR and (b) we estimate $H(n,\omega)$ rather than the TFR. Such a transfer function is used not only to obtain positive estimates of the deblurred TFR using (2) but also for signal reconstruction, masking and filtering of non–stationary signals, and in the modeling of time–varying systems.

The rest of the paper is organized as follows. In Section 2, we discuss the estimation of the generalized transfer function via deconvolution. We show that the deconvolution as stated above admits many solutions for $H(n,\omega)$. In Section 3, we present some applications of the new estimate of the GTF.

2. ESTIMATING THE GENERALIZED TRANSFER FUNCTION

In this section we present a deconvolution technique to estimate the generalized transfer function. We first consider an unconstrained deconvolution and observe that, while the resulting estimate of $H(n,\omega)$ has some interesting properties, it does not display the change in frequency of a signal with time. To overcome this problem we perform a constrained deconvolution.

2.1. Unconstrained Deconvolution

To estimate $H(n,\omega)$, we minimize the following mean squared error function

$$MSE_1 = \sum_{n=0}^{N-1} \int_{-\pi}^{\pi} |E\{S_{BD}(n,\omega)\} - \beta_1 S_{BDH}(n,\omega)|^2 \, d\omega \quad (6)$$

where β_1 is a normalization factor,

$$S_{BDH}(n,\omega) = \int_{-\pi}^{\pi} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} W[l-n,k] H^*(l-k,\lambda) H(l+k,\lambda) e^{-j(\omega-\lambda)2k} d\lambda \quad (7)$$

is the right hand side of (4) and N is the length of the signal being considered. Since in practice only one realization of the signal and consequently its TFR is available, we use $S_{BD}(n,\omega)$ instead of $E\{S_{BD}(n,\omega)\}$.

The error measure in (6) has no unique minimum. In fact, any function of the form

$$H(n,\omega) = F(\omega) x(n) e^{-j\omega n} \quad (8)$$

where $F(\omega)$ is an arbitrary, unit–energy function of ω, yields a zero MSE. This is easily shown by replacing (8) in (7).

It is interesting to remark that the signal can be recovered from any of the solutions given by (8). In fact we have that

$$\int_{-\pi}^{\pi} H(n,\omega) e^{j\omega n} d\omega = x(n) \int_{-\pi}^{\pi} F(\omega) d\omega = x(n) \quad (9)$$

Furthermore, the resulting ES is positive and satisfies the time marginal [4], i.e.,

$$\int_{-\pi}^{\pi} |F(\omega) x(n) e^{-j\omega n}|^2 d\omega = |x(n)|^2.$$

Finally, if $F(\omega) = X(\omega)$, the Fourier transform of $x(n)$, the resulting ES also satisfies the frequency marginal [4] up to a scale factor

$$\sum_{n=0}^{N-1} |X(\omega) x(n) e^{-j\omega n}|^2 = |X(\omega)|^2 \sum_{n=0}^{N-1} |x(n)|^2.$$

Unfortunately the expression in (8) results in a separable function of time and frequency. Such a representation only describes tones (or impulses) well. It is important to note that many solutions (not in form of (8)) which are not separable in time and frequency exist. To obtain such solutions, we need to impose additional constraints on the minimization problem.

2.2. Constrained Deconvolution

The objective is to obtain an $H(n,\omega)$ that minimizes the MSE in (6) and displays changing frequencies with time. To this end, we formulate a constraint using the Wigner distribution (WD). Namely, we minimize the MSE in (6) and at the same time constrain the resulting ES to be as close as possible to the WD without cross terms. For this we define a constraining error measure

$$MSE_2 = \sum_{n=0}^{N-1} \int_{0}^{\pi} \left| S_{WD}(n,\omega) - \beta_2 |H(n,\omega)|^2 \right|^2 M(n,\omega) d\omega \quad (10)$$

where β_2 is a normalization factor and $M(n,\omega)$ is a mask function used to suppress the cross terms. The mask function is set to the lowest entropy evolutionary periodogram (EP) [3, 7]. This substantially reduces the effect of the cross terms of the Wigner distribution.

The constrained deconvolution is then carried out by minimizing the following cost function

$$c = MSE_1 + MSE_2 \quad (11)$$

where MSE_1 is the original error measure given by (6) and MSE_2 is the constraining error function given by (10). The two error measures in (11) can be minimized separately. Possible solutions for MSE_1 are of the form given by (8), whereas the solution for MSE_2 is given by

$$|H(n,\omega)| = \begin{cases} \sqrt{S_{WD}(n,\omega)} & S_{WD}(n,\omega) > 0, \\ 0 & \text{otherwise,} \end{cases} \quad (12)$$

with arbitrary phase. However, none of these solutions minimize both errors at the same time and thus are not minimum solutions of the cost function in (11). Since, it is not possible to obtain a closed form solution to this cost function, we solve the problem iteratively using the conjugate gradient method. As an initial estimate of the GTF, we use the one obtained from the minimum entropy EP [3, 7].

The resulting GTF provides a positive estimate of the ES with lower entropy than the initial estimate. Compared with the EP, this GTF estimated spectrum is not limited by the order of the orthonormal expansion functions and as such it does not have the time–frequency trade–off of the EP. Furthermore, when blurred according to (3) TFRs similar to those computed directly from the signal can be obtained.

Just as in the unconstrained deconvolution case, the estimated GTF can be used to reconstruct the signal. While the reconstruction is not perfect, it still offers a good approximation. Furthermore, the estimate of the evolutionary spectrum obtained from the constrained deconvolution satisfies the time marginal under a special condition. Namely, if MSE_1 given by (6) is reduced to zero, and the Wigner distribution is chosen as the TFR to deblur.

3. APPLICATIONS OF THE GTF

3.1. Spectral Representation of Speech Signal

We apply the constrained deconvolution technique to a 31.25 millisec speech signal sampled at 8kHz (a portion of the utterance "..... credit card, and we use that", spoken by a female) [6]. The signal contains a sweeping formant and some stationary formants. We used the GTF from EP with $M = 5$ as our initial estimate and its magnitude square as the mask function. The ES estimate corresponding to the deblured GTF is shown in Fig. 1 and has an entropy of 13.5 bits. For comparison, Fig. 2 shows an estimate of the TFR of the speech signal obtained by combining narrow–band, wide–band and medium–band spectrograms. It has an entropy of 14.4 bits. From the figures it is clear that while the spectrogram estimates suffer from excessive blurring and are not able to track the frequency modulation of the sweeping formant, the estimate of the ES obtained by solving the constrained deconvolution clearly shows the sweeping formant and simultaneously resolves the stationary formants. Furthermore, the ES obtained from the new estimate of the GTF has a lower entropy than the one obtained from a combination of spectrograms [6] showing that the former is more informative than the latter.

3.2. Time–Frequency Masking of the GTF

Time–frequency masking [2, 12] has been used to obtain signals corresponding to a particular region in the time–frequency plane. Time–frequency masking was recently applied [11] to estimates of the GTF. The method consists of multiplying the estimate of the GTF of the signal $x(n)$ by a positive mask function $G(n, \omega)$ to obtain the GTF of the desired signal $y(n)$. The desired signal $y(n)$ is then synthesized through the Wold-Cramer relation given by (1). The mask function is constructed such that it is equal to one in the desired regions of interest and zero otherwise.

Now as an example consider a two–chirp signal given by

$$x(n) = \frac{n(N-n)}{N^2} e^{j(\frac{\pi}{2} - \frac{\pi n}{4N})n} + 0.22 e^{j(\pi - \frac{4\pi n}{9N})n} \qquad (13)$$

where the length of data N used was 128. We choose to deblur the EP to obtain an estimate of the GTF by solving the constrained deconvolution problem described in Section 2.2. The ES corresponding to this estimate of the GTF is depicted in Fig. 3. A mask function along the chirp rate of the first component of the signal given by (13) is specified, i.e., the one whose frequency changes from $\frac{\pi}{2}$ to 0 as n goes from 0 to $N-1$. This mask function is shown in Fig. 4. We multiply this mask function with the above mentioned GTF to obtain a modified GTF. Using this modified GTF we obtain a signal that is plotted as dashed lines in Fig. 5. Also shown as solid line in the same figure is the actual signal, i.e., the first component of the signal in (13). As can be seen the recovery is almost perfect except for the small error at the ends. The EP estimate of the signal displayed in Fig. 6 shows that this signal corresponds to the correct time–frequency region as specified by the mask function.

4. CONCLUSIONS

We presented a deconvolution technique to estimate the GTF of an LTV system. Deconvolution has been used before in time–frequency but the main purpose has been to obtain a deblurred TFR. Estimation of the GTF not only permits us to obtain a deblurred TFR but has many additional advantages as demonstrated by the applications presented. We presented constraints that permit us to arrive at a desirable answer and illustrated the advantages of the new method with the help of examples.

REFERENCES

[1] Special Issue on Time–Frequency Analysis, *Proceeding of the IEEE*, Vol. 84, No. 9, September 1996.

[2] M. R. Portnoff, "Time-Frequency Representations of Digital Signals and Systems Based on Short-Time Fourier Analysis", *IEEE Trans. on Acoust., Speech and Signal Proc.*, Vol. ASSP-28 No. 1, pp. 55–69, 1980.

[3] A. S. Kayhan, A. El-Jaroudi and L. F. Chaparro, "Evolutionary Periodogram for Non-Stationary Signals", *IEEE Trans. on Signal Proc.*, Vol. 42 No. 6, pp. 1527–36, June 1994.

[4] L. Cohen, *Time–Frequency Analysis*, Prentice-Hall, NJ, 1995.

[5] L. White, "Resolution Enhancement in time-frequency signal processing using inverse methods", *IEEE Proc. ICASSP'89*, pp. 1622–1625, 1989.

[6] J. W. Pitton, L. E. Atlas and P. J. Loughlin, "Application of Positive Time-Frequency Distribution to Speech Processing", *IEEE Trans. on Speech and Audio Proc.*, Vol. 2 No. 4, October 1994.

[7] S. I. Shah, P. J. Loughlin, L. F. Chaparro and A.El-jaroudi, "Informative Priors for Minimum Cross–Entropy Positive Time Frequency Distributions", *IEEE Signal Processing Letters*, Vol. 6, No. 4, pp. 176-177, June 1997.

[8] G. Melard and A. H. Schutter, "Contributions to Evolutionary Spectral Theory", *J. of Time Series Analysis*, Vol. 10 No. 1, pp. 41–63, January 1989.

[9] M. B. Priestley, *Spectral Analysis and Time Series*, Academic Press Inc., 1981.

[10] C. S. Detka, A. El-Jaroudi and L.F. Chaparro, "Relating the Bilinear Distributions and the Evolutionary Spectrum", *Proc. of ICASSP'93, Vol. IV, pp. 496-499*, 1993.

[11] A. Akan and L. F. Chaparro, "Evolutionary Spectral Analysis Using a Warped Gabor Expansion", Signal Processing, accepted for Publication.

[12] W. Krattenthaler and F. Hlawatsch, "Time-Frequency Design and Processing of Signals Via Smooth Wigner Distributions", *IEEE Trans. on Signal Processing*, Vol. 41 No. 1, pp. 278–287, January 1993.

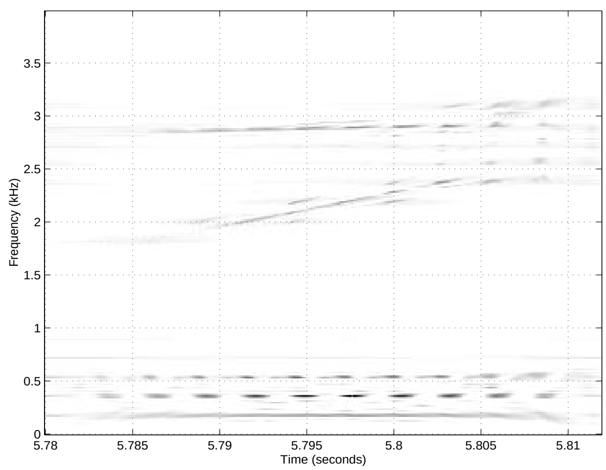

Figure 1. ES of the speech signal

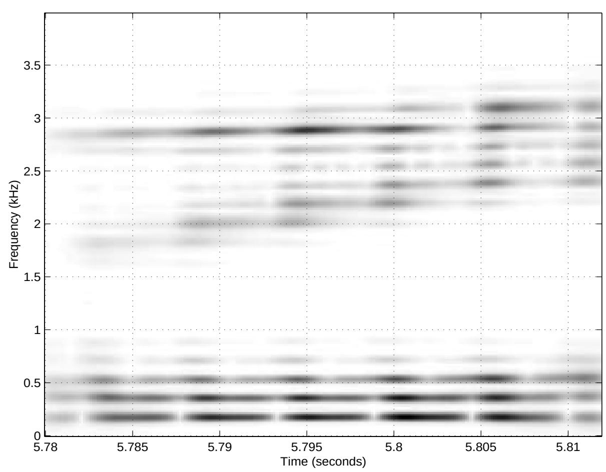

Figure 2. Combination of three spectrograms: speech signal

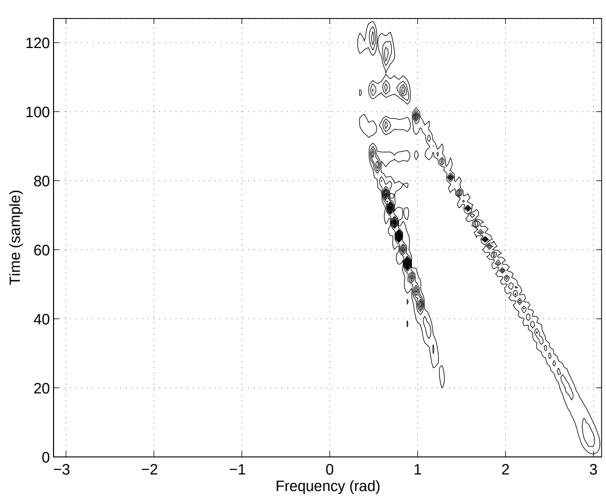

Figure 3. ES of the two chirp signal

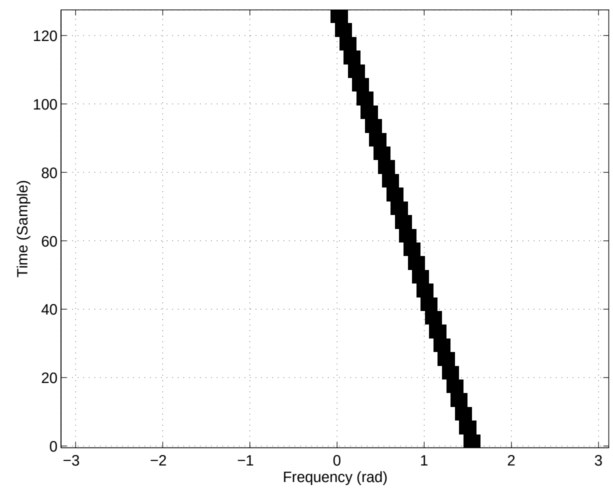

Figure 4. Mask Function

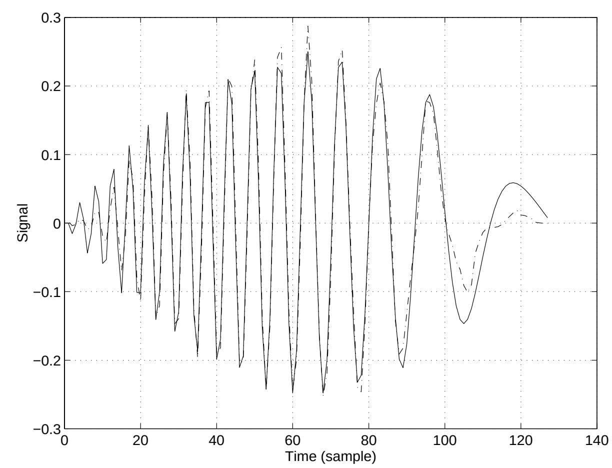

Figure 5. Reconstruction, solid: actual, dashed: estimate

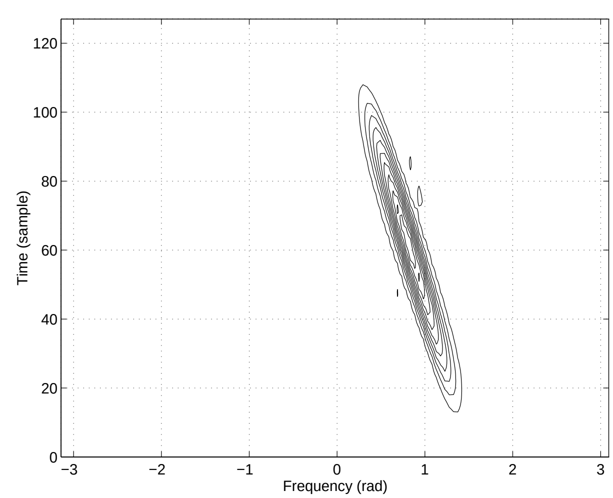

Figure 6. EP of the signal obtained using the modified GTF