DYNAMIC TRACKING FILTERS FOR DECOMPOSING NONSTATIONARY SINUSOIDAL SIGNALS

Ashwin Rao *Ramdas Kumaresan*

Department of Electrical Engineering
The University of Rhode Island
Kingston RI 02881

ABSTRACT

A procedure to decompose a signal consisting of nonstationary sinusoidal components based on the principles of Residual Signal Analysis [1, 2] is proposed. A tracking unit consisting of an all zero filter (AZF) in cascade with a dynamic tracking filter (DTF) is assigned to each component. While the adaptively varying zeros of the AZF suppresses all interfering neighbors, the DTF captures the slowly varying instantaneous frequency (IF) of the desired component. Our earlier methods improved upon Costas's estimator-predictor filter bank [2] by using a better interfering signal predictor. The alternative procedure described in this paper avoids the use of prediction and is not overly restricted by the number of components. We also show that by using two simple feedback loops (a loop-filter as in [2] is thus avoided) the tracking information is ensured to be in phase. Finally, the algorithms ability to decompose synthetic signals, speech signals into harmonic partials as well as tracking the formants present in voiced speech segments is illustrated.

1. INTRODUCTION

The problem of tracking multiple nonstationary sinusoids present in a given composite signal finds applications in areas such as co-channel signal separation, interfering signal cancellation as well as speech analysis. We model a signal $x[n]$ consisting M nonstationary components as

$$x[n] = \sum_{k=1}^{M} a_k[n] \exp(j\phi_k[n]) + w[n] \qquad n = 0, 1, 2, \ldots \tag{1}$$

where $w[n]$ is additive white Gaussian noise, $\phi_k[n]$ is the phase track of the k-th component and $a_k[n]$ is its envelope track. The IF of the k-th component is defined as $f_k[n] = (\phi_k[n] - \phi_k[n-1])/2\pi$. We can think of the above signal $x[n]$ as having M sinusoids, each with its own meandering frequency tracks and slowly varying envelope tracks. Our aim is to acquire and accurately track each of these components. In this paper we address tracking only.

A traditional fixed filter-bank cannot isolate the components because the frequency of each component is sweeping, at times rapidly, which would cause it to appear in the passband of different filters at different times. In parametric modeling techniques, the model parameters are estimated by minimizing the squared error between the signal and its model over a certain time interval; the averaging interval is thus the same for for all components. In practice, the individual components are generally accompanied with envelopes and IFs that vary at different rates. In that case, ideally, the averaging intervals should be different for the different components. But the modeling techniques work with the composite data, in which all underlying components get treated identically irrespective of the rate of change of their envelopes and frequencies. Many adaptive filtering techniques like the resonator-based filter-banks, the adaptive line enhancer (ALE) and the adaptive notch filter have been proposed [7–9]. In these methods the frequency estimates that adapt the filter parameters are estimated from the composite signal; usually a gradient descent procedure is used. However, when the components have varied strengths, a stronger component might bias the estimate of a relatively weaker one. This in turn could result in inaccurate tracking.

Residual Signal Analysis (RSA), originally proposed by Costas [1], utilizes an estimator-predictor filter bank to achieve signal decomposition. The key idea is to treat all neighboring components as interference while tracking a desired component. The basic difference between RSA and the adaptive filtering techniques is that in RSA all interfering components are coherently subtracted to form a residual signal; the tracking information, which is the frequency of the desired signal, is then estimated from this residue. Interference suppression is achieved by predicting the next time-sample of all neighbors from past estimates and then subtracting them from the composite signal. To improve the suppression of interfering components, in our previous procedure [2] called Residual Interfering Signal Canceler (RISC), the group delay was compensated by solving a set of first-order linear predictor equations in a least squares sense. However, its accuracy decreases as the number of predictor equations increase.

In this paper we use what is called a dynamic tracking filter (DTF) [4] to track the IF of a nonstationary component. The primary advantage is its ability to successfully track real-valued signals without explicitly computing the quadrature components [3], as opposed to the time-varying demodulation technique of RSA. Using multiple DTFs, we show how to decompose accurately a multicomponent sig-

This work was supported by the Air Force Contract AFOSR F49620-93-I-0026 and by The Center for Research On Applied Signal Processing at The University of Southern California.

0-7803-2431-5/95 $4.00 © 1995 IEEE

nal. The need for a predictor is avoided by inserting an all zero filter, ahead of the DTF, whose zeros are continuously adjusted. Further, to compensate the delay introduced in the estimator loop, we introduce an additional loop: the input to this second DTF is delayed by an amount corresponding to the group delay of the lowpass filter within the first loop.

2. DYNAMIC TRACKING FILTER (DTF)

Dynamic Tracking Filters are single resonance filters which are designed to follow the changing frequency of a signal. In a DTF, the input signal passes through a narrow filter followed by an IF estimator. The estimated IF of the signal is then fed back to adapt the tracking filter's pole location. This causes the DTF to follow the IF of the incoming sinusoidal component. Thus the dynamic tracking filter is just a time-varying bandpass filter which can be realized by the simple difference equation

$$y[n] = r_p \exp\left(j2\pi \tilde{f}_p[n]\right) y[n-1] + (1-r_p)x[n] \quad , \qquad (2)$$

where $x[n]$ is the input to the DTF and $y[n]$ is its output. r_p is the radius of its single pole and $\tilde{f}_p[n]$ is the estimated IF of $x[n]$. The constant $1-r_p$ ensures unity gain at the resonant frequency of the DTF. In general, the IF may be estimated using any one of a large number of methods including Prony's method. We estimate the frequency by taking the difference in angle of the neighboring samples of the filter output; details are in [2].

2.1. DUAL-LOOP DTF FOR IN-PHASE TRACKING

Consider the feed-back loop shown in Figure 1 that estimates the IF needed to adapt the tracking filter. In the presence of noise and interference, the phase derivative will also be noisy. A lowpass filter (LPF) is therefore used following the angle differencer, to capture the signal IF. But this results in an estimate that is delayed; the delay corresponds to the group delay of the averaging filter. To circumvent this problem, we delay the input by an equivalent amount and feed it to another DTF. In the area of frequency demodulation, the use of such dual-loops was one of the early solutions to reduce the phase lag introduced in feedback loops of tracking circuits. Note that the second DTF *does not* have an explicit feed back loop of its own; its resonant frequency location is the IF estimated by the feed back path of the first DTF.

3. ALGORITHM FOR TRACKING MULTIPLE NONSTATIONARY COMPONENTS

When there are multiple components, the k-th DTF should track only the k-th component. For this to happen, the k-th AZF's output should contain only the k-th component. The remaining components can be rejected if the AZF nulls are placed at their frequencies. The information about where to place the nulls is derived from the other DTFs. For example, the center frequency information of the DTF tracking the l-th component is used to place a zero at that location in the k-th AZF. If this is done for all the remaining $M-1$ components, then the k-th AZF's output will contain only the k-th component, and that DTF will behave (ideally) as though it is tracking a single component.

Figure 1 shows the block diagram of one of the channels of the filter bank. The box labeled AZF is the adaptive all zero filter whose $M-1$ zeros are at the estimated IFs of all its $M-1$ neighbors. The transfer function of the AZF of the k-th tracker at any time index n is

$$H_{Ak}(n,z) = K[n] \times \prod_{\substack{l=1 \\ l\neq k}}^{M} (1 - r_z e^{j2\pi \tilde{f}_l[n]} z^{-1}) \qquad (3)$$

where

$$K[n] = \frac{1}{\prod_{\substack{l=1 \\ l\neq k}}^{M} \left(1 - r_z e^{j2\pi(\tilde{f}_l[n] - \tilde{f}_k[n])}\right)} \qquad (4)$$

ensures unity gain and zero phase lag at the estimated IF of the kth component. The DTF tracking the k-th component has a transfer function

$$H_{Dk}(n,z) = \frac{1-r_p}{1 - r_p e^{j2\pi \tilde{f}_k[n]} z^{-1}} \quad . \qquad (5)$$

r_z is the radius of the zero (≈ 0.99), r_p is the pole radius and $f_k[n]$ is the IF of the kth component at time index n.

The IF estimates are the smoothed outputs of the angle differencer. To avoid problems due to phase unwrapping the angle differencer is implemented as $\angle\,(y_k[n-1]y_k^*[n-2])$ [6], where $*$ denotes complex conjugation. An example of a filter transfer function used to smooth the IF estimate is

$$H_l(z) = \frac{(1-r_l)^2}{(1-r_l z^{-1})^2} \quad , \qquad (6)$$

where the choice of the pole-radius r_l depends on the rate of change of the component's envelope and IF. Finally, observe that the envelopes of the separated components are the outputs ($\hat{x}_k[n]$ in Figure 2) of a second set of DTFs.

4. COMPUTER SIMULATIONS

The above proposed algorithm was first tested on synthetic data. The reader is referred to [2] for details on the generated signals. The composite signal consisting of four components (sampling frequency 4 kHz) was fed to the DTF filter bank shown in Figure 1. The pole locations of the four DTFs were initialized at the true IF values for the first 5 ms. All DTFs had a bandwidth of 210 Hz (pole radius 0.92) and each LPF was an IIR filter having a double pole of radius 0.9 at origin. The estimated frequency tracks are plotted in Figure 2(a) as solid lines; dashed lines correspond to the true ones. In Figure 2(b) we have plotted the envelopes estimated by the second loop DTFs; the AZFs in the second loop were supplemented by Costas's predictor (optional) for cleaner envelope estimates. A single realization of zero mean complex white Gaussian noise with variance $\sigma^2 = 200$ was then added to the same composite signal. The maximum signal-to-noise ratio (SNR$_{max}$) was ≈ 20.5 dB while SNR$_{min}$ was ≈ -10.5 dB. SNR$_{max}$ and SNR$_{min}$ were defined as $20\log_{10}\frac{A_{max}}{\sqrt{\sigma^2}}$ and $20\log_{10}\frac{A_{min}}{\sqrt{\sigma^2}}$

respectively; $A_{max} = 150$ and $A_{min} = 4.5$ being maximum and minimum of the envelopes. Our algorithm was then tested on this noisy signal; it was switched to tracking after 30 ms. Figure 2(c) displays the narrowband spectogram along with the true (dashed lines) and the estimated (solid lines) IF tracks. We next applied our method decompose real speech data.

The speech signal was taken from the TIMIT database (sample number 4351 to 8100 of sentence `si1111.wav` sampled at 16 kHz). It was filtered by a LPF with order 300 and 3 dB cutoff 1.8 kHz and decimated to 4 kHz to reduce computation. An analytic signal was then formed by using Hilbert transformation. Again, all the DTFs and the LPFs had pole radii of 0.92 and 0.9 respectively. For the first 30 ms, the resonant frequency track of the k-th DTF was initialized by k times the fundamental frequency. The separated *nominally* harmonic partials are plotted in Figure 3(a). Notice that the algorithm cleanly separated the first seven harmonic tracks; observe the departure of some higher partials from strict harmonicity. The envelopes of these partials are plotted in Figure 3(b).

We now show that by increasing the bandwidths of the DTFs, the same algorithm can track formants present in speech. A synthetic formant was first generated as described in [3]. The signal was pre-emphasized by a $1 - 0.95z^{-1}$ filter. A single DTF with pole radius 0.98 was used to track the formant; LPF used had pole radius 0.986. Figure 4 shows the spectogram as well as the true (dashed) and the estimated (solid line) formant tracks. Next, the speech segment was fed to a bank of four DTFs. Our aim was to simultaneously track the pitch and the first three formant regions. The algorithm parameters were as follows: First four DTFs were initialized (for 5 msecs) at 250 Hz, 500 Hz, 2000 Hz and 2800 Hz. They had pole radii $0.96, 0.92, 0.9$, and 0.88 respectively. All LPFs had a second-order pole (radius 0.94) at d.c. We have plotted the resonance frequency tracks (solid thick lines in Figure 5) of the DTFs which seem to closely match with the pitch and the three formant regions observed in the wideband spectogram. Figure 6 shows the analysis results of a *male* voiced speech; portion "oily" from TIMIT sentence /timit/test/dr2/mcem0/sa2.wav. Finally, it is possible to extract the higher formants by suitably bandpass filtering or pre-emphasizing the data prior to analysis.

5. CONCLUSION

We showed that multiple dynamic tracking filters, when supplemented by adaptive all zero filters, can be used in a filter bank structure for accurately decomposing multicomponent nonstationery signals. The proposed method avoids the use of prediction-subtraction to form residual signals as in [2]. Accurate decomposition of clean as well as noisy synthetic signals was demonstrated. It was shown that our procedure can be applied for retrieving the harmonic partials present in voiced speech signals as well as for simultaneous pitch and formant tracking.

6. REFERENCES

[1] J.P. Costas, *Proc. of First ASSP Workshop on Spectral Estimation*, Hamilton, Canada, Aug'81, pp. 6.5.1–6.5.8.

[2] R. Kumaresan, C.S. Ramalingam and A. Rao, *Proc. of Seventh SSAP Workshop*, Quèbec City, Canada, Jun. 1994, pp. 207–210.

[3] C.S. Ramalingam, A. Rao and R. Kumaresan, *Proc. of IEEE-SP Symposium on Time-Frequency and Time-Scale Analysis*, pp 500–503, Oct. 1994, PA,USA.

[4] J. H. Roberts, *Proc. of IEE*, vol 115, No 11, November 1968

[5] V. Bykov, *Telecomm. Radio Engg.*,Pt. I, 18, 1964, pp. 17–26.

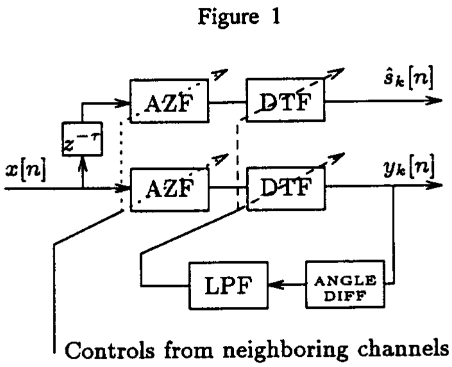
[6] S. M. Kay, *IEEE Trans. Acoust., Speech and Signal Proc.*, ASSP-37, 1989, pp. 1987–1990.

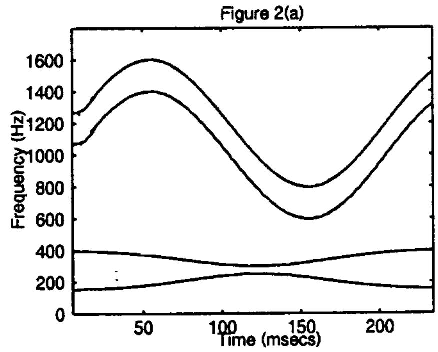
[7] B. Widrow et al., *Proc. IEEE*, vol. 63, pp. 1692–1716, Dec. 1975

[8] D. V. B. Rao and S. Kung, *IEEE Trans. Acoust., Speech and Signal Proc.*, Vol. ASSP-32,No. 4, August 1984.

[9] M. Padmanabhan and K. Martin, *IEEE Trans. Circuits and Sys.*, vol. 38, No. 10, oct. 1991.

Acknowledgments: The authors wish to thank Mr. C.S. Ramalingam for many useful discussions.





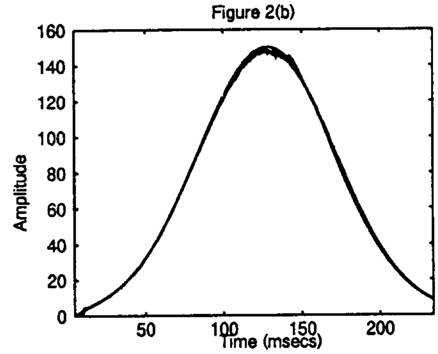
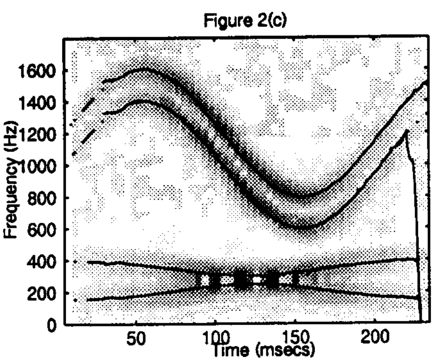
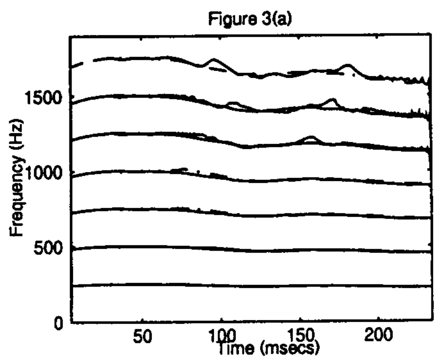


Fig. 1 The k-th channel in the DTF bank used for multicomponent signal decomposition. The all zero filter (AZF) receives information (from the LPF outputs in other channels) about the IFs of all neighboring components. By adaptively placing zeros at those frequencies it suppresses all interfering components; it ensures that the input to the k-th channel DTF is approximately the k-th signal component. **Fig. 2**(a) Estimated IF tracks (solid lines) and True IF tracks (dashed lines) of a 4 component synthetic signal (no noise) match exactly. **Fig. 2**(b) True (dashed lines) and Estimated (solid lines) envelope tracks.





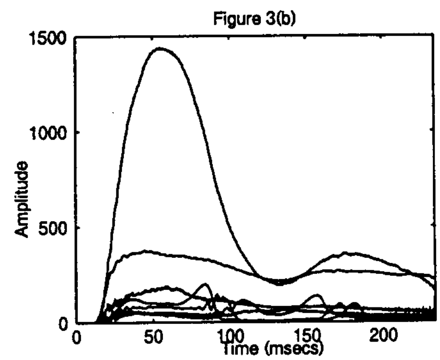
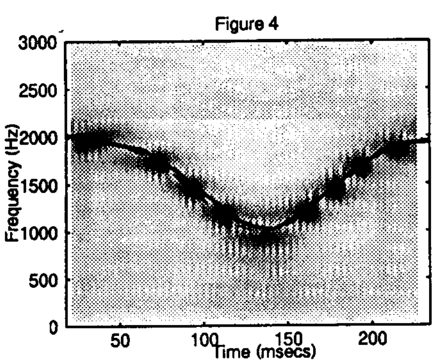
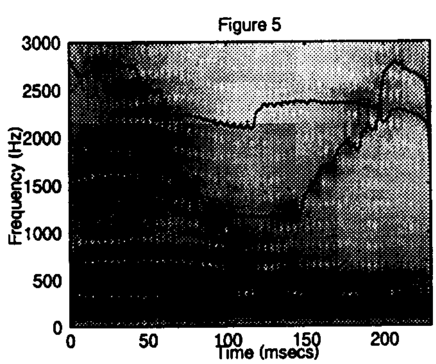


Fig. 2(c) The Spectogram, True (dashed lines) and Estimated (solid lines) IFs of a noisy version (additive white Gaussian noise) of the same 4 component AM-FM signal; SNR_{max} and SNR_{min} were ≈ 20.5 & -10.5 dB respectively. **Fig. 3**(a) Nominally harmonic tracks (solid lines) separated from a voiced speech segment: 'si1111.wav' from TIMIT database. k-th dashed lines represents $k \times$ the fundamental frequency; observe the departure from strict harmonicity displayed by the high frequency partials. **Fig. 3**(b) Estimated envelope tracks of the seven harmonics displayed in Fig 3(a).





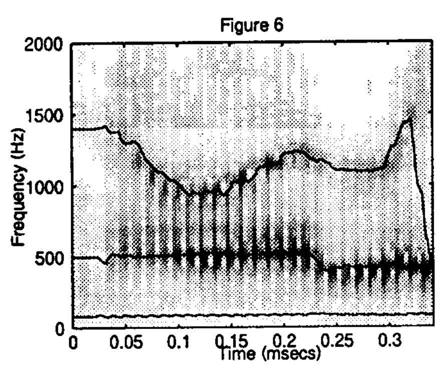


Fig. 4 A synthetic formant was generated by exciting a time-varying single resonant filter by synthetic glottal pulses. The Spectogram along with the True (dashed) and Estimated (thick line) formant tracks are plotted. **Fig. 5** The Spectogram, the Estimated pitch ($\approx$ 250 Hz) and the formant tracks (both displayed by solid lines) of the speech segment used in Fig. 3. **Fig. 6** The Estimated pitch ($\approx$ 90 Hz) and two formants in *male* voiced speech: "oily" corresponding to `/timit/test/dr2/mcem0/sa2.wav` in timit database.