26th IEEE Southeastern Symposium on System Theory (SSST'94), Athens, OH, pp. 208-212, March 20-22, 1994.

Parametric Modeling and Reconstruction of Acoustic Signals[1]

John F. Tilki and A.A. (Louis) Beex

The Bradley Department of Electrical Engineering
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061-0111

Abstract

We report on the development and performance evaluation of parametric signal processing algorithms for extracting a mixed stochastic signal process from its wideband noise corrupted measurements. The signals under consideration are acoustic with periodic components masked by wideband colored noise.

First, the unbiased autocorrelation function (ACF) sequence is estimated from the data. The higher lags of the ACF are used in a generalized version of the least-squares modified Yule-Walker equation estimator to obtain accurate sinusoidal frequency estimates. Then the relative sinusoidal amplitudes and phases are found by maximum likelihood estimation or by modal decomposition. Once the sinusoids have been completely characterized, the wideband components of the signal are modeled using AR or ARMA spectral estimation procedures.

1. Introduction

We want to accurately characterize, model, and reconstruct acoustic signals. The signals being considered consist of periodic components masked by wideband colored noise. Typical household engines, motors, and machinery are sources of such signals, and the wideband noise is often created by turbulent air flow in the system. Wideband noise can also be introduced by the measurement process itself, and by other background noise from the environment.

The sound data to be analyzed is sampled at 44.1 kHz and quantized to 16 bits of resolution. Since ample data is available, we can expect that accurate autocorrelation function (ACF) estimates can be made, even at fairly low SNR. Therefore, we use an unbiased ACF estimator as the basis for many of the analysis procedures.

First, high lags of the ACF are used in a least-squares modified Yule-Walker equation estimation procedure (LSMYWE) [1, 2] to obtain sinusoidal frequency estimates. Given the frequencies, maximum likelihood estimation [3] is used to estimate the corresponding amplitudes and phases. Alternately, modal decomposition can be employed to estimate the sinusoidal amplitudes [4]. After the sinusoids have been completely characterized, the residual wideband component can be represented by an AR (or ARMA) model.

2. Generalized LSMYWE

Complex sinusoids can be modeled by an AR process with poles approaching the unit circle as the variance of the driving white noise approaches zero [3, 4]. Although the wideband component of the signal masks the narrowband components at lower ACF lags, at high lags the ACF becomes almost purely periodic. Therefore we expect to extract sinusoidal component information from an AR model based on the higher ACF lags.

A unique relationship exists between the parameters of an ARMA system and its autocorrelation sequence. Given an ARMA(p,q) linear shift invariant system

$$H(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1 z^{-1} + \cdots + b_q z^{-q}}{1 + a_1 z^{-1} + \cdots + a_p z^{-p}} \qquad (1)$$

with a driving white noise input u[n], a linear difference equation relates the output x[n] to u[n]

$$x[n] = -\sum_{k=1}^{p} a[k]x[n-k] + \sum_{k=0}^{q} b[k]u[n-k] \qquad (2)$$

[1] Supported by IBM Federal Systems Division, Manassas VA, under Grant #WM-138704, and the Center for Innovative Technology, Herndon VA, under Grant #INF-93-008.

A similar relation exists for the output correlation $r_{xx}[k]$ in terms of the ARMA parameters

$$r_{xx}[k]=\begin{cases}-\sum_{l=1}^{p}a[l]r_{xx}[k-l]+\sigma_u^2\sum_{l=0}^{q}b[l]h^*[l-k] & ;k\in[0,q]\\ -\sum_{l=1}^{p}a[l]r_{xx}[k-l] & ;k\geq q+1\end{cases} \quad (3)$$

The MA influence on the ACF expires after the q-th lag. Therefore the AR parameters of an ARMA process may be found by solving a set of p or more linear equations which result from (3) for $k > q$ with $M\text{-}l \geq p$

$$\begin{bmatrix} r_{xx}[l+1]\\ r_{xx}[l+2]\\ \vdots\\ \vdots\\ r_{xx}[M]\end{bmatrix}=-\begin{bmatrix} r_{xx}[l] & r_{xx}[l-1] & \cdots & r_{xx}[l-p+1]\\ r_{xx}[l+1] & r_{xx}[l] & \cdots & r_{xx}[l-p+2]\\ \vdots & \vdots & \vdots & \vdots\\ \vdots & \vdots & \vdots & \vdots\\ r_{xx}[M-1] & r_{xx}[M-2] & \cdots & r_{xx}[M-p]\end{bmatrix}\begin{bmatrix} a[1]\\ a[2]\\ \vdots\\ a[p]\end{bmatrix} \quad (4a)$$

($l > q$), or in matrix form

$$\underline{r}=-R\underline{a}. \quad (4b)$$

When the theoretical ACF is replaced by an estimate, as is the case in practice, the equations are no longer satisfied. To account for estimation errors, (4b) becomes

$$\underline{\hat{r}}=-\hat{R}\underline{a}+\underline{e} \quad (5)$$

where $\underline{\hat{r}},\hat{R}$ denote estimates, and the error term $\underline{e}$ is zero if the estimates equal the theoretical values. The least-squares estimate of the AR parameters is then

$$\underline{\hat{a}}=-(\hat{R}^H\hat{R})^{-1}\hat{R}^H\underline{\hat{r}} \quad (6)$$

The pole locations of this AR model are calculated, and any poles which occur within a narrow band around the unit circle are classified as sinusoidal. The angles of these poles constitute our frequency estimates.

3. Sinusoidal parameter MLE

Although accurate sinusoidal frequency estimates are available from the LSMYWE, the relative amplitudes and phases of the sinusoids have not yet been determined. In order to calculate the sinusoidal amplitudes and phases, a maximum likelihood estimation (MLE) approach [3] is applied assuming that the sinusoidal frequencies are known.

Suppose that the sound data consists of p complex sinusoids in complex white Gaussian noise

$$x[n]=\sum_{i=1}^{p}A_i e^{j(2\pi f_i n+\phi_i)}+z[n] \quad ;\ n\in[0,N-1] \quad (7)$$

where $z[n]$ is the complex white Gaussian noise with zero mean and variance σ_z^2. The sinusoidal frequencies f_i are assumed to be known, but the amplitudes and phases, A_i and ϕ_i respectively, are to be estimated. To better formulate the problem, first represent the sinusoidal amplitudes and phases as a single complex amplitude

$$A_{ci}=A_i e^{j\phi_i} \quad (8)$$

Then define A_c and E as follows

$$\mathrm{A}_c=\begin{bmatrix}A_{c1} & A_{c2} & \cdots & A_{cp}\end{bmatrix}^T \quad (9)$$

$$\mathrm{E}=\begin{bmatrix}e_1 & e_2 & \cdots & e_p\end{bmatrix} \quad (10)$$

where

$$e_i=\begin{bmatrix}1 & e^{j2\pi f_i} & \cdots & e^{j2\pi f_i(N-1)}\end{bmatrix}^T \quad (11)$$

With the sinusoidal frequencies known, or replaced by their maximum likelihood estimates, the maximum likelihood estimates of the sinusoidal amplitudes and phases are

$$\mathrm{A}_c=\left(\mathrm{E}^H\mathrm{E}\right)^{-1}\mathrm{E}^H x \quad (12)$$

where

$$x=\begin{bmatrix}x[0] & x[1] & \cdots & x[N-1]\end{bmatrix}^T \quad (13)$$

4. Modal decomposition

The fact that the human ear is insensitive to audio phase distortion is well known. To verify this assumption in the context of this project, the authors conducted several audio tests with human observers. In all cases the human observers were unable to discern phase distortions in acoustic signals of the type being considered. Therefore, it can be concluded that accurate sinusoidal phase estimates are not essential if the quality of signal reconstruction is to be measured psychoacoustically (i.e., by human observers).

When sinusoidal phase information is not relevant and can be arbitrarily assigned, the relative sinusoidal amplitudes can alternately be calculated from a modal decomposition of the ACF sequence [4]

$$r_{xx}[k] = \sum_{i=1}^{p} A_i \rho_i^k \tag{14}$$

where $r_{xx}[k]$ is the k-th lag, the A_i are complex modal amplitudes, and the ρ_i are the p modes (poles) of the system. Using the system poles found via the LSMYWE, along with the ACF estimates, we can solve for the modal amplitudes by forming the l equations ($l \geq p$)

$$\begin{bmatrix} r_{xx}[k] \\ r_{xx}[k+1] \\ \vdots \\ \vdots \\ r_{xx}[k+l-1] \end{bmatrix} = \begin{bmatrix} \rho_1^k & \rho_2^k & \cdots & \rho_p^k \\ \rho_1^{k+1} & \rho_2^{k+1} & \cdots & \rho_p^{k+1} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \rho_1^{k+l-1} & \rho_2^{k+l-1} & \cdots & \rho_p^{k+l-1} \end{bmatrix} \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_p \end{bmatrix} \tag{15a}$$

or in matrix form

$$\underline{r} = P\,\underline{A} \tag{15b}$$

An overdetermined system of equations can be solved in a least-squares sense in order to reduce the estimator variance. Thus the estimates of the modal amplitudes are

$$\underline{\hat{A}} = -(\hat{P}^H \hat{P})^{-1} \hat{P}^H \underline{\hat{r}} \tag{16}$$

5. Signal reconstruction

Once the sinusoidal components have been completely characterized, the wideband colored noise remains to be modeled. This can be done by a number of conventional AR (or ARMA) spectral analysis procedures, such as the autocorrelation method [5].

When the entire signal model is complete, the signal may be reconstructed by a simple superposition of its components. All sinusoids with their proper amplitudes and phases are summed. White noise is filtered through the AR model of the wideband component, and added to the sinusoidal sum, yielding the complete signal reconstruction.

6. Experimental Results

Experimental results on signals measured from a kitchen blender indicate that while the wideband component masks the narrowband components at lower ACF lags, at high lags the ACF becomes almost purely periodic (see Figure 1). The latter implies that the LSMYWE can be expected to perform well as a frequency estimator.

In order to evaluate the performance of the procedures, a known test signal was created with characteristics similar to the acoustic signals being considered. The test signal consists of the first ten harmonics of a square wave with a normalized fundamental frequency of 0.02, immersed in Gaussian white noise (see Figure 2). The variance of the noise was chosen such that the SNR for the fundamental component is 0 dB.

The unbiased ACF was estimated from 1 second of sound data (44,100 data points) and 2,000 lags of the ACF beginning at lag 50 were passed to the LSMYWE routine. An AR model order of 35 was specified (2 poles for each of the 10 real sinusoidal harmonics, and arbitrarily 15 poles were chosen to model the noise). The pole locations of the resulting AR model were calculated, and any poles with radii in [0.98, 1.02] were classified as sinusoidal. These sinusoidal pole frequencies are listed in Table 1, and the corresponding pole diagram of the entire LSMYWE AR model is provided in Figure 3. Note that the model poles occur on the unit circle at precisely the sinusoidal harmonic frequencies. This is true even for the weaker harmonics which are about 25 dB below the noise level. Thus the modified LSMYWE can be expected to perform well as a sinusoidal frequency estimator when applied to similar acoustic signals.

The fact that the sinusoidal poles fall on the unit circle, and all "extra" poles model the wideband component, suggests that perfect AR model order selection is not critical, and this procedure will perform well given a wide range of AR model orders above some minimum. This hypothesis was supported experimentally.

The results of next applying the sinusoidal parameter MLE to 8,000 data points of the square wave test signal are also given in Table 1. The parameter estimates for the stronger harmonics are quite accurate, and the estimates for the weaker harmonics are reasonably accurate, given their SNR of -20 dB or less. The output signal reconstructed from the MLE model closely resembles the approximate square wave of the test signal (see Figures 4 and 5).

Finally modal decomposition was applied to the test signal ACF, and the results are listed in Table 1. An order *p* of 35 was specified, and 350 equations were used. Given the low SNR of the weaker harmonics, the sinusoidal amplitude estimates are again reasonably accurate. It should be noted that both the MLE and the modal decomposition estimates are accurate to within computer precision if no noise is present in the signal.

The MLE and modal decomposition sinusoidal models were each used to reconstruct audio signals. When compared to the original test signal (no noise) in repeated audio tests, no differences could be perceived. This was true when the original/reconstructed combination was played in immediate sequence, as well as if there was a time delay between the two.

Next the autocorrelation method was applied to the test signal, and an AR(6) model was found for the wideband noise component. This wideband model was then used in conjunction with the MLE and modal decomposition sinusoidal models in an attempt to reconstruct the entire test signal. When audio tests were performed, most subjects could perceive little or no difference between the original and reconstructed signals if a time delay of a few seconds separated the playback of the two signals. However, when the two signals were played in immediate succession, a small difference could be perceived. This difference seemed to exist as a low frequency noise presence. Plots of the spectra of the reconstructed signals revealed that indeed the wideband noise model possessed too much energy at low frequencies. These low frequency mounds in the noise spectra are due to using the entire test signal, including the sinusoids, in the wideband AR modeling procedure. Hence the AR wideband model catches some of the sinusoidal power at the stronger harmonics. Methods are currently being developed to eliminate such sinusoidal influence on the wideband modeling process.

7. Conclusions

The algorithms presented in this paper have proven useful in characterizing and modeling a mixed stochastic process. The above version of the LSMYWE can provide accurate sinusoidal frequency estimates if the process ACF can be estimated reliably. Once accurate frequency estimates are available, the maximum likelihood estimation procedure can yield sinusoidal amplitude and phase estimates. If phase information is not relevant, modal decomposition gives an alternative approach to amplitude estimation. Both methods performed well despite the low signal-to-noise ratios in the test signal.

REFERENCES

[1] J. Cadzow, "High Performance Spectral Estimation – A New ARMA Method," IEEE Trans. ASSP-28, pp. 524-529, Oct. 1980.

[2] A. A. (Louis) Beex and Louis L. Scharf, "Covariance Sequence Approximation for Recursive Digital Filter Design," *ICASSP-80*, pp. 260-263, Apr. 9-11, 1980.

[3] Steven M. Kay, Modern Spectral Estimation, Theory & Application (Englewood Cliffs, New Jersey: Prentice Hall, 1988).

[4] A. A. (Louis) Beex and Louis L. Scharf, "Covariance Sequence Approximation for Parametric Spectrum Modeling," *IEEE Trans. ASSP-29*, No. 5, pp. 1042-1052, Oct. 1981.

[5] J. Makhoul, "Linear Prediction: A Tutorial Review," *Proc. IEEE*, Vol. 63, pp. 561-580, Apr. 1975.

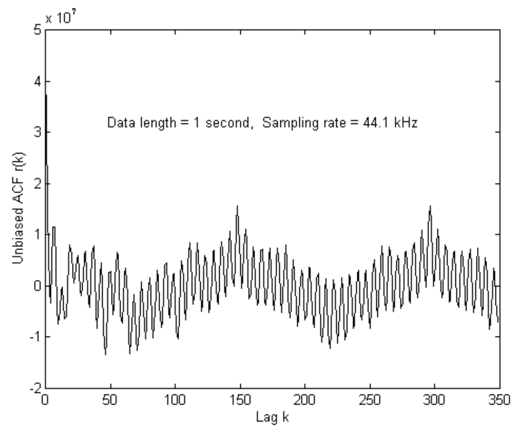

Figure 1: ACF of Kitchen Blender

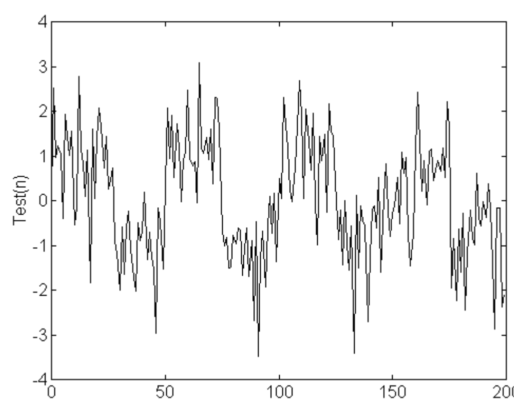

Figure 2: Test Signal with Noise

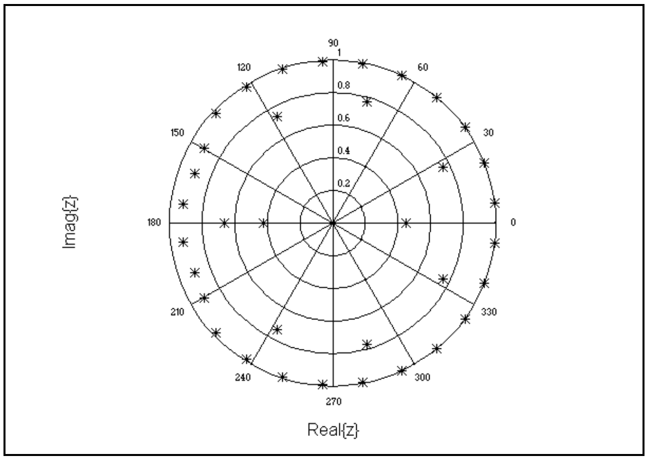

Figure 3: LSMYWE AR Model Pole Locations

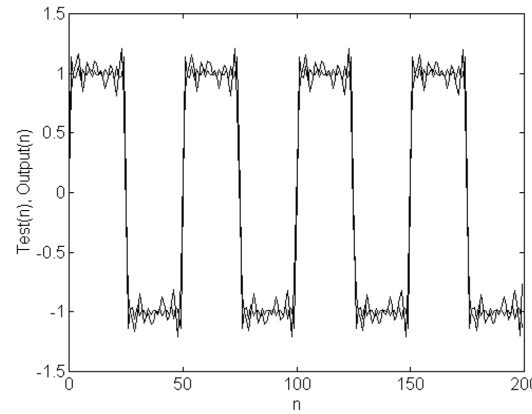

Figure 4: Test Signal and Reconstruction

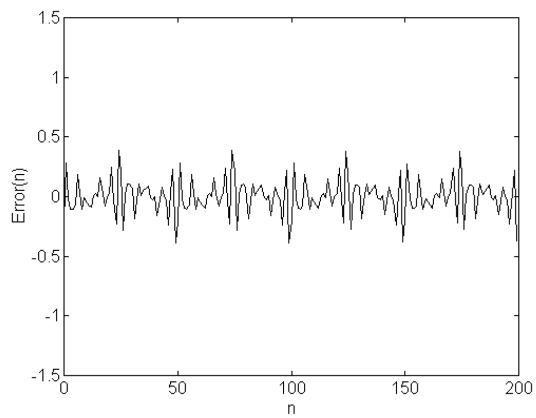

Figure 5: Error Signal

Test Signal Sinusoidal Parameters				LSMYWE Estimates	Sinusoidal MLE Estimates		Modal Decomposition Estimates
Frequencies (normalized)	Amplitudes	Phases (radians)	Harmonic SNR (dB)	Frequencies (normalized)	Amplitudes	Phases (radians)	Amplitudes
0.0200	0.6366	0.0000	0.0000	0.0200	0.6376	0.0333	0.6379
0.0600	0.2122	0.0000	-9.5424	0.0600	0.2011	-0.0023	0.2138
0.1000	0.1273	0.0000	-13.9794	0.1000	0.1310	0.1086	0.1254
0.1400	0.0909	0.0000	-16.9020	0.1400	0.0838	-0.1914	0.0912
0.1800	0.0707	0.0000	-19.0849	0.1800	0.0611	0.0655	0.0799
0.2200	0.0579	0.0000	-20.8279	0.2200	0.0460	0.9740	0.0766
0.2600	0.0490	0.0000	-22.2789	0.2599	0.0415	1.2557	0.0719
0.3000	0.0424	0.0000	-23.5218	0.2999	0.0181	0.1440	0.0741
0.3400	0.0374	0.0000	-24.6090	0.3398	0.0065	2.3263	0.0874
0.3800	0.0335	0.0000	-25.5751	0.3798	0.0079	3.9161	0.0779

Table 1: Original and Estimated Sinusoidal Parameters