

# YASAS - YET ANOTHER SOUND ANALYSIS-SYNTHESIS METHOD

*Shlomo Dubnov*

Department of Music, University of California, San Diego

## ABSTRACT

This paper describes a novel analysis-synthesis method based on estimation and comparison between spectral methods that are specifically designed for modeling of random signals. These methods are an alternative to the discrete Fourier transform, providing high-resolution, smooth and easily interpretable spectral estimation over short signal frames. It is generally acknowledged that musical signals that exhibit both periodicity and variations can be modeled as sinusoidal components with various extents of modulation (generally characterized as band-limited frequency components), and remaining noise or wide-band frequency components that describe unvoiced and other noisy parts of the signal spectrum. Using single frame Fourier analysis makes it difficult to tell if spectral energy at a particular frequency is due to noise or sinusoidal components. Existing methods for such decomposition usually consider properties of Fourier phases, which are noisy. Our method decomposes a signal into sinusoidal, modulated sinusoidal and noise components based on a comparison between two spectral representations, namely autoregressive (AR) and minimum variance distortionless response (MVDR), both calculated from linear prediction coefficients (LPC) in a time varying manner. Using different optimal properties of each model, we develop estimators for frequencies and amplitudes (spectral envelope) of sinusoidal components (spectral lines) in noise and derive a “noisality” index that assigns different weights to contributions of sinusoidal and noise components at every frequency. Examples of synthetic and real sounds are presented in the paper.

*Keywords*— **Sound Analysis-Synthesis, Spectral Estimation, LPC, AR, MVDR, Sinusoidal and Noise Model, Modulated Periodicity, Bandwidth-Enhanced Additive Model**

## I. INTRODUCTION

Sound analysis-synthesis methods are required for a variety of applications, including sound editing and transformation, modeling of acoustical sources, such as musical instruments and voices, sound compression, transmission and more. These methods are based on

a detailed analysis of the time varying frequency properties of a sound, representing the signal in terms of changing frequency contents and identifying its basic sources of variation. One of the most popular modeling techniques for musical signals is sinusoidal analysis [1][2]. These methods decompose a sound into collection of sinusoids with time varying amplitudes, frequencies and sometimes also phases. Such decomposition is usually done using Fourier analysis over short time frames, followed by steps of peak detection, interpolation and tracking of partials over sequence of FFT vectors. Due to the finite frequency resolution of the resulting short time FFT analysis, interpolation is required in order to find the exact frequencies of the spectral peaks. Moreover, modeling of noisy or modulated parts of the signal is possible by sinusoids with fast and random varying phase if the analysis times step is sufficiently small (few milliseconds). When longer analysis steps are used, or when phases are not estimated, it is necessary to decide at the analysis step whether an energy peak at a particular frequency belongs to sinusoidal or noise sources, in order to resynthesize them accordingly. The heuristics for sinusoidal versus noise decomposition include checking for existence or lack of particular phase relations among adjacent frequency bins [3] and/or phase continuity across FFT bins in subsequent frames [4]. More elaborate methods employ additional steps for separate modeling noisy parts of the signal, such as obtaining a residual part by careful subtraction of the sinusoidal components from the original signal [5], or using random phase modulations for broadening of the spectral peaks at the sinusoidal frequencies [6][7]. Mixed harmonic and noise [8] and harmonic, sinusoidal and noise [9] methods appeared also in speech and general audio modeling applications and are known for their favorable properties in terms of quality versus synthesis versatility and compression tradeoffs.

In this paper we develop a new method for analysis of sounds that is completely different from existing sinusoidal methods in the sense that it does not rely upon standard Fourier analysis. We use methods of parametric spectral estimation [10], namely autoregressive (AR) and minimum variance distortionless re-

sponse (MVDR), where the parameters of both models are derived from a single Linear Prediction (LP) [11] estimate that is performed in time varying fashion over successive frames containing short segments of signal samples. The AR model assumes that the analyzed signal results from passing a white noise through a resonant all pole filter. AR can be also used to represent narrow spectral lines caused by sinusoids, which are modeled as very narrow resonances (poles on the unit circle) at precise frequencies of the sinusoids. One of the main advantages of using AR for estimation of spectral lines is that it is not limited in terms of frequency resolution, giving precise frequencies of spectral lines even for short segments of the signal. It is known, though, that AR modeling of spectral lines can be severely degraded if the signal is corrupted by noise. This limitation severely limits the applicability of AR modeling in the case of mixed sinusoidal and noisy signals, like the music signals that we are investigating. We show that this limitation can be overcome by using a high order AR model, which is later reduced based on comparison to MVDR spectrum. The MVDR spectral analysis method has almost complementary properties: it is derived from signal powers at the output of an especially designed filter bank that passes a signal at particular frequency in a distortionless manner, with minimum interference by sinusoidal components at other frequencies. This results in spectral estimates that have optimal estimates of sinusoidal signal powers at different frequencies that are statistically robust and have a smooth shape across frequencies, but also having poor frequency resolution that does not allow for detection of sinusoidal peaks when the frequency of the sinusoidal components is unknown a-priori. Using a comparison between AR and MVDR estimate we are able to detect the spectral peaks with high precision and also derive a “noisality” index, that is used for performing separate resynthesis of the sinusoidal and noise components.

## II. BACKGROUND ON PARAMETRIC SPECTRAL ANALYSIS

In this section we present a short overview of the AR and MVDR spectral analysis methods and present a simple iterative algorithm that relates the parameters of the two representations.

### II-A. AR spectral Analysis

The AR model is defined by

$$x(n) = \sum_{i=1}^p a_i x(n-i) + e(n) \quad (1)$$

where  $x(n)$  is the current value of the time series,  $a_1, \dots, a_p$  are predictor coefficients,  $p$  is the model order (also indicating the number of the past values used to predict the current value) and  $e(n)$  represents a one-step prediction error, i.e. the difference between the predicted value and the current value at this point. The AR model can be used to create two filters: a prediction error filter through which  $x(n)$  is filtered to produce the error sequence  $e(n)$  and a synthesis filter that recreates the samples  $x(n)$  by filtering  $e(n)$  as its input. The optimum predictor coefficients are such that the prediction error is minimized. There are efficient methods for computation of optimal predictor coefficients from data samples, generally known as linear prediction [11].

It should be noted that common use of LP is for the purpose of modeling spectral envelopes in speech signals that roughly correspond to resonant properties of the human vocal tract. These models are low order and are not intended for capturing individual sinusoidal components. AR model can be used for spectral analysis by estimating the optimal filter and using the spectral shape of the synthesis filter to describe the signal spectrum. This spectrum shows how the power (variance) is distributed as a function of frequency once the oscillatory components in the time series are produced by a filter that spectrally shapes or colors a white input signal (the error signal).

Writing the AR equation in the spectral domain (using Z transform for the sequences  $x(n)$  and  $e(n)$ , giving  $X(z)$  and  $E(z)$  respectively) we have  $X(z) = H(z)E(z)$ , where

$$H(z) = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2} - \dots - a_p z^{-p}} \quad (2)$$

The estimates for the different frequency components of a time series can be calculated from the poles of  $H(z)$ , i.e., the roots of the denominator of the  $H(z)$  polynomial.

The spectrum of the modeled time series,  $P_{AR}(e^{j\omega})$  (also to be denoted as  $P_{AR}(\omega)$ ), is obtained by multiplying the prediction error variance  $\sigma_e^2$  with the square of the transfer function.

$$P_{AR}(e^{j\omega}) = \sigma_e^2 |H(e^{j\omega})|^2 = \frac{\sigma_e^2}{|1 - a_1 e^{-j\omega} - \dots - a_p e^{-jp\omega}|^2} \quad (3)$$

### II-B. The MVDR Spectrum

The Minimum Variance Distortionless Response (MVDR) spectrum is a flexible spectral analysis method which was introduced by Capon [12], and is mostly employed in array processing applications, and has also been investigated in relation to other applications such as speech modeling [13], audio filtering

[14] and voicing detection [15]. The MVDR spectrum  $P_{MV}(\omega)$  is given by

$$P_{MV}(\omega) = \frac{1}{\mathbf{v}^H(\omega)\mathbf{R}_{p+1}^{-1}\mathbf{v}(\omega)} \quad (4)$$

where  $\mathbf{v}(\omega) = [1 \ e^{j\omega} \ e^{j2\omega} \ \dots \ e^{-jp\omega}]$  is a frequency tuning vector with  $\mathbf{v}^H$  denoting its conjugate transpose, and  $\mathbf{R}_{p+1}$  is autocorrelation matrix. The spectrum order  $p$  corresponds to the largest correlation lag in the autocorrelation matrix.

The MVDR spectrum has an affinity with non-parametric filterbank spectral analysis methods. In the filter-bank interpretation of periodogram (FFT based) spectral analysis methods, the spectrum at any given frequency can be viewed as the power at the output of a bandpass filter. In this case the bank of bandpass filters is data independent and its characteristics are defined by the length and choice of the analysis window, arranged along an equally spaced frequency grid. Moreover, the frequency characteristics of the individual filters are such that they are also frequency independent between themselves. Similarly to periodogram-based methods, the MVDR spectral estimate can be conceptually viewed as an output of a bank of filters, with each filter centered at one of the analysis frequencies. But, in contrast to FFT based methods, the bandpass filters of the MVDR bank are both data and frequency dependent, an information that is captured by the signal autocorrelation matrix  $\mathbf{R}_{p+1}$  appearing in the definition of the MVDR spectrum in equation (4). In particular, the MVDR spectral estimate at frequency  $\omega$  utilizes a specially designed FIR bandpass filter with impulse response  $\mathbf{w} = \{w(n), n = 0..p\}$  whose passband characteristics are designed so that it would pass distortionlessly the signal components at the frequency  $\omega$ , expressed mathematically as

$$W(\omega) = \sum_{i=0}^p w(i)e^{-jk\omega} = \mathbf{w}\mathbf{v} = 1. \quad (5)$$

and also minimizing the overall filter output power for a given input signal, expressed as  $\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R} \mathbf{w}$ . These filter design and error minimization steps are done in the course of deriving the MVDR spectrum. Details of this derivation are not brought here due to space limitations. The MVDR spectrum is a set of output powers of these filters, which were designed according to the desired passband and attenuation characteristics as described above, and specifically for the analyzed signal. In other words, the MVDR filter designed for frequency  $\omega$  will let the input signal component at frequency  $\omega$  pass through undistorted (passband), and will minimize the total output power of the filter, a situation

that ensures that the remaining frequency components of that specific signal are suppressed in an optimal manner by putting deep nulls (notches) at competing frequencies existing in the specific signal. This process is repeated for every frequency component, giving the complete MVDR spectrum.

## II-C. Relation between AR and MVDR spectra

The MVDR parameters can be related to the AR spectrum in a very interesting and computationally efficient way. The AR spectrum is parameterized by a set of Linear Prediction Coefficients (LPC)  $a_k, k = 1..p$ , (we also include  $a_0 = 1$  in the equations below) and a prediction error variance (gain)  $\sigma_e^2$ . The parameters  $\mu(k)$  of the MVDR spectrum can be directly obtained using a non-iterative computation based upon the LPC [16][10]

$$\mu(k) = \begin{cases} \frac{1}{\sigma_e^2} \sum_{i=0}^{p-k} (p+1-k-2i)a_i a_{i+k}^* & k \geq 0 \\ \mu^*(-k) & \text{otherwise} \end{cases} \quad (6)$$

Using this parameterization, the MVDR spectrum can be written as

$$P_{MV}(\omega) = \frac{1}{\sum_{k=-p}^p \mu(k)e^{-jk\omega}}. \quad (7)$$

Thus, in order to obtain both AR and MVDR spectra, it is sufficient to perform a single LPC analysis of the signal samples in every frame and transform the AR  $a_i$  coefficients into  $\mu(k)$  coefficients of MVDR representation. It should be noted that this method does not actually design and implements the MVDR filters that could be used to filter a signal in a distortionless manner at every frequency. What the current method does is allow estimating the MVDR spectrum for all frequencies, “as if” the signal was passed through a bank of MVDR filters where every filter was designed for the particular signal at every frequency, and then the power at the output of this filter bank was recorded.

It was shown by Burg [17] that the MVDR spectrum can be written as the harmonic mean of all AR spectra for orders 0 to  $p$ :

$$\frac{1}{P_{MV}(\omega)} = \sum_{k=0}^p \frac{1}{P_{AR(k)}(\omega)}, \quad (8)$$

where we introduce the notation  $AR(k)$  to denote the order of the AR model. As a result of this averaging, the MVDR spectrum has a smoother appearance and lower resolution than the corresponding  $p$ -th order AR spectrum.

### III. MODEL OVERFITTING AND REDUCTION

The LP method [11] attempts to reduce the overall variance of the prediction residual by trying to match the AR synthesis filter to the signal spectrum. This can be viewed as an attempt of the LP method to create a prediction error filter that places zeros on the poles of the signal spectrum, thus reducing to minimum the filter output and minimizing the prediction error. For a single sinusoid, it is evident that LP can predict the signal with zero error, expressing a sinusoid at frequency  $\omega_0$  as the following AR equation<sup>1</sup>

$$x(n) = 2\cos(\omega_0)x(n-1) - x(n-2) \quad (9)$$

This is a two pole filter, with poles located at angles  $\omega_0$  and  $-\omega_0$  on the unit circle. When the signal consists of few sinusoidal partials only, the LP poles move closer to the unit circle as the filter order is raised, until the number of poles is twice the number of sinusoids (the factor two is due to prediction of a real signal, which requires two poles at conjugate locations for every sinusoidal component). It can be shown that in case of  $L$  sinusoidal components, AR model of order  $p > 2L$  can perfectly fit the signal, resulting in zero error.

#### III-A. Estimation of sinusoids in noise

In the case when both sinusoidal and noise component are present in the signal, the LP method drastically changes its performance. In order to understand this effect let us consider a signal modeling problem where we observe a signal that is a sum of two processes, one sinusoidal and the other colored noise. Each individual signal can be efficiently represented by an AR filter: In the case of sinusoidal signal, the signal is represented by an AR model with poles close to (practically “on”) the unit circle. In such a case, the prediction can be exact, resulting in residual error that is very close to zero. Let us assume that the noise part could be written in terms of a white noise passing through an AR filter. In this case the poles are inside the circle closer to the origin, and the variance of the excitation white noise, which equals to the prediction residual, is much higher. When we sum both components, the resulting spectrum, by construction, is a sum of two finite order AR processes. This can be written as a sum of two AR filters, resulting in a filter that can be written as a ratio of two polynomials (called “rational filter”), where the polynomials in the denominator and numerator are of order higher than one. To demonstrate this

<sup>1</sup>The relation is simpler for a complex phasor  $e^{-j\omega_0 n}$ , resulting in order 1 AR filter.

effect, let us assume filter

$$B(z) = \frac{\sigma}{1 - 2R\cos(\theta)z^{-1} + R^2z^{-2}} \quad (10)$$

that represents the noise, with pole at radius  $R$  at angle  $\theta$  and filter

$$A(z) = \frac{g}{1 - 2r\cos(\omega)z^{-1} + r^2z^{-2}} \quad (11)$$

that represents a sinusoidal component (the gain  $g$  is used to scale the amplitude of the narrow band noise that approximates a sinusoid). Note that we have moved the pole of the sinusoidal model just a little bit inside the unit circle by assuming a pole with radius  $r < 1$ ,  $r \rightarrow 1$ , to avoid singularities. The sum filter becomes

$$C(z) = A(z) + B(z) = \frac{g(1 - 2R\cos(\theta)z^{-1} + R^2z^{-2}) + \sigma(1 - 2r\cos(\omega)z^{-1} + r^2z^{-2})}{(1 - 2r\cos(\omega)z^{-1} + r^2z^{-2})(1 - 2R\cos(\theta)z^{-1} + R^2z^{-2})}, \quad (12)$$

which is a rational polynomial with nominator of second degree and denominator of fourth degree. This filter, and accordingly its power spectrum, have both poles and zeros. The poles remain at the original locations, but the zeros add spectral notches at new locations, which are the roots of the new nominator polynomial. The resulting spectrum is no more a pure AR but is actually a combination of autoregressive and a moving average (MA) process, called ARMA.

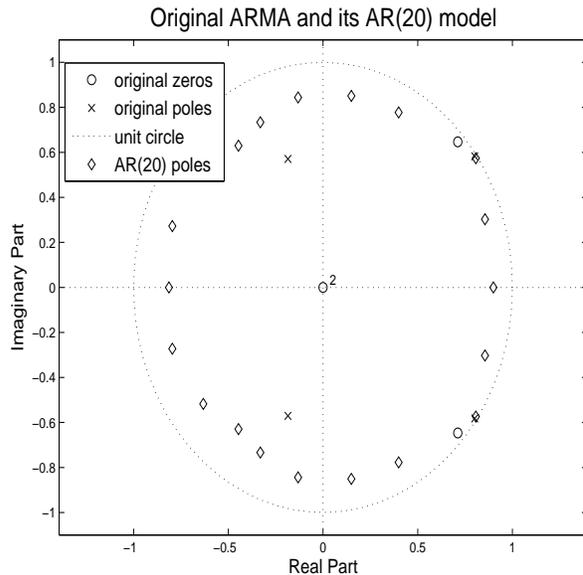


Fig. 1. Pole zero plot of mixed signal and its AR approximation.

It should be noted that any finite ARMA process can be represented by an infinite order AR model. This can

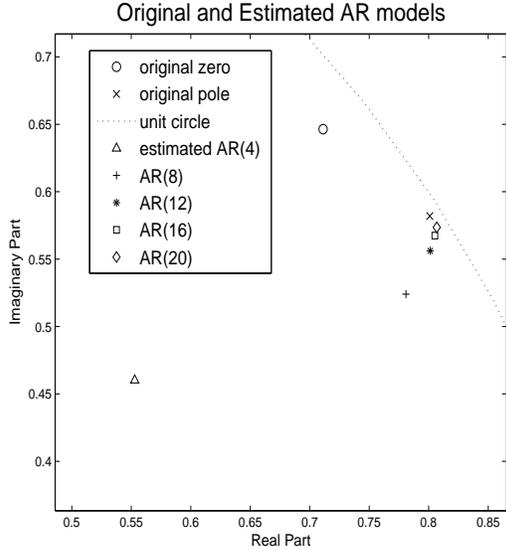


Fig. 2. Progressive change in pole locations near the sinusoidal pole for increasing orders of AR model.

be shown by performing a polynomial division of the AR part by the MA filter. The residual polynomial after division can be expanded into an infinite series, resulting in a high order AR model. When ARMA signal is modeled by low order AR, the poles in the vicinity of the sinusoidal component are drawn closer to origin in direction opposite from location of the zero. As the order of AR increases, the pole matching the sinusoidal part approaches its true location. The additional poles spread almost evenly around  $2\pi$ , overall matching the remaining noise part of the spectrum.

Figure 2 shows the results of increasing the order  $p$  of an estimated AR model, denoted by  $AR(p)$ , for a signal comprising of a sum of narrow band (i.e. approximately sinusoidal) and wide band signals. Both signals are represented by  $AR(2)$  models, the narrow band signal with two conjugate poles of radius 0.99 at angles  $\pm 0.2\pi$  and a noise signal with poles of radius 0.6 at angles  $\pm 0.6\pi$ . The summation creates additional zeros of radius 0.961 at angles  $\pm 0.235\pi$ , resulting in an ARMA model.

Figure 3 shows a comparison between power spectrum estimate of a signal created by passing a white noise through ARMA filter and the spectrum of an  $AR(20)$  model estimated for the same signal. The original ARMA filter is plotted in dashed line. It can be seen that  $AR(20)$  matches rather correctly the sinusoidal (narrowband) component, while the wide-band part is captured in a "wavy" manner by multiple poles.

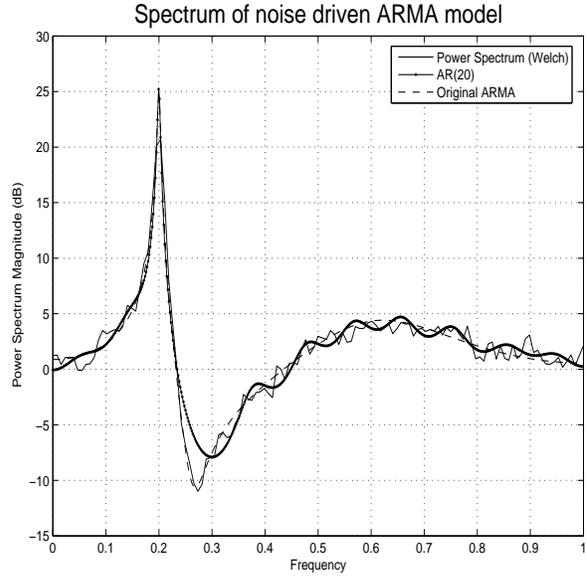


Fig. 3. Welch and  $AR(20)$  power spectrum estimates of noise driven ARMA model

### III-B. Sinusoidal extraction and parameter estimation

As explained above, the MVDR spectrum gives an estimate of the power of the signal at every frequency in terms of the power of the output of a bandpass filter that is optimally designed, both data and frequency wise, so that a sinusoidal frequency at every frequency passes through undistorted. It was shown above the sufficiently high order AR spectrum is capable of modeling signals that are mixture of sinusoidal and noise components, placing the poles of the AR model "on" the sinusoids and "spreading" the remaining poles over the remaining noise envelope. If the AR peak significantly exceed the power of MVDR estimate, it is plausible to assume that a sinusoidal component is present at this frequency. If the powers of both spectral estimators are comparable, there is little evidence for presence of a periodic component. The ratio of the spectral powers of the two spectral methods can be used to separate the AR model into sinusoidal and noise parts, or could be used continuously as a "noisality" measure to weight noise and sinusoidal contributions at every partials.

Figure 4 shows analysis of one frame from a recording of voice signal from a Suzanne Vega song. The graph show LPC, MVDR and standard FFT (Periodogram) spectra. Visual inspection of the graph reveals six frequency peaks whose LPC spectrum is significantly higher than MVDR. These partial components appear at frequencies 213.89, 416.86, 1691.10, 1908.46, 2546.92, 2749.49 Hz, which are almost har-

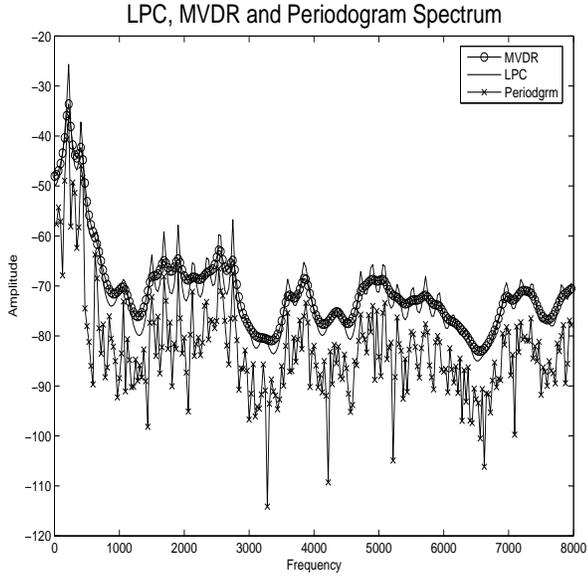


Fig. 4. Comparison between different types of spectra in a single frame.

monically related to the lowest frequency of 213.89 Hz, with multiplicative factors of 1.94, 7.90, 8.92, 11.91, 12.85, i.e. approximately partials 2,8,9,12,13, with the highest partial deviating by 15% from the fundamental.

#### IV. RESULTS

We present results of analysis - synthesis of a vocal solo singing recording by Suzanne Vega. The analysis was performed with AR(80), analysis frame of 24 msec. and overlap 50%. In order to achieve partial tracking, a simple best winner continuation selection was applied for every pole, with maximum allowed deviation in frequency of approximately a semitone (6%) and maximal allowed change in amplitude not exceeding 12 dB. Figure 5 shows the time varying frequencies of the different poles of Vega singing signal. In order to distinguish between the sinusoidal and noise parts, the analysis parameters were divided into two groups based on noisality, with time-frequency points having AR to MVDR power ratio above 4.2dB being classified as sinusoidal, and those below this ratio classified as noise.

Making hard decision between sinusoidal and noise components is undesirable since there is not clear distinction between various levels of sinusoidality and noisiness and as can be seen from the figure, there are continuous transitions between noise and sinusoids over time. Accordingly, instead of generating two separate signals, we used all poles to generate both sinusoidal and noise signals, and applied changing relative weights to sinusoidal and noise parts according to the following

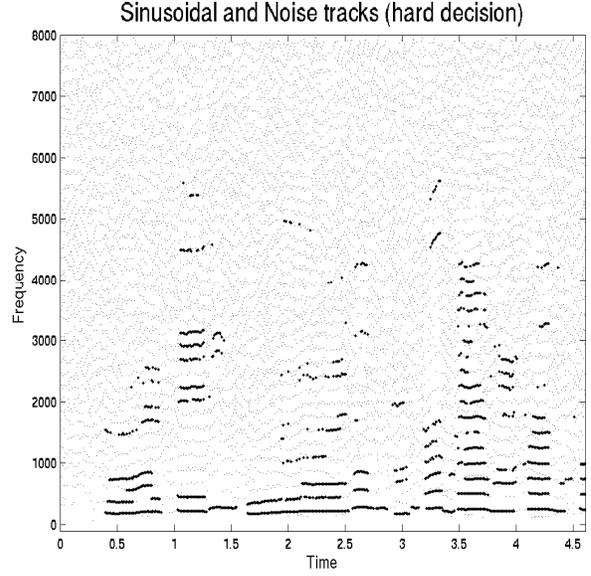


Fig. 5. Sinusoidal and noise track of Vega singing.

relation:

$$x_r(t) = \sum_{k=1}^p [\gamma_k(t)s_k(t) + \sqrt{1 - \gamma_k^2(t)}n_k(t)] \quad (13)$$

where  $x_r(t)$  is the reconstructed signal,  $s_k(t)$  and  $n_k(t)$  are sinusoidal and noise signals of the  $k$ -th track, generated according to MVDR amplitude at frequencies of the poles for the sinusoidal part, and according to AR filter model for the noise part, with  $\gamma_k(t)$  being a time varying weight (“noisality index”) for track  $k$ , whose value at time  $t$  are derived according to the AR to MVDR power ratio. So, for frequencies whose AR spectrum significantly “overshoots” the MVDR spectrum,  $\gamma$  is set close to 1, while for cases that MVDR is higher than AR spectrum, the noise model dominates by setting *gamma* close to zero. The values of  $\gamma_k(t)$  implement this approximate mapping using a sigmoid function that is calibrated to  $\gamma$  value 0.5 at power AR to MVDR power ratio of 6dB, and range of  $\approx \pm 5$ dB from zero to unit  $\gamma$  values.

Additionally, sinusoidal “widening”, or band-limited modulation of sinusoidal components can be approximately estimated in terms of the pole radius. When the radius  $r$  is close to 1, the 3-dB bandwidth  $B$  (in Hertz) can be approximated by  $B \approx -\frac{\ln(r)}{\pi T}$ , where  $T$  is the sampling interval (in seconds). This allows introducing modulation with the appropriate bandwidth into the sinusoidal components, resulting in “bandwidth enhanced” sinusoids. It should be noted that bandwidth estimation is very difficult in FFT based models, since it is limited by the analysis rate for determining phase

or instantaneous frequency values for consequent estimation of their deviations. Other criteria, such as matching the complex FFT or effectively comparing the theoretical complex analysis window to the actual FFT is related to noisality in non-trivial manner and requires sophisticated and empirically calibrated threshold for deciding between sinusoidal and noise bands [3]. Other statistical models [18] based on fluctuations of the temporal envelope are reported at this point for synthetic signals only.

In the resynthesis, sinusoidal frequency changes over time are made continuous by interpolation between frequencies and amplitudes across consecutive frames in every track. Synthesis of the noise part does not require detailed interpolation. Moreover, in order to avoid excessive accumulation of noise from separate noise components at different frequencies, each noise component is band-limited. The bandwidth of the noise component is chosen to be proportional to signal sampling frequency divided by the model order. The intuition behind this limit is that the noise part is approximately described by equally spread poles around the  $2\pi$  range.

The following figures 6 - 9 show spectrograms of the original, noise, sinusoidal and combined (sum) signal.

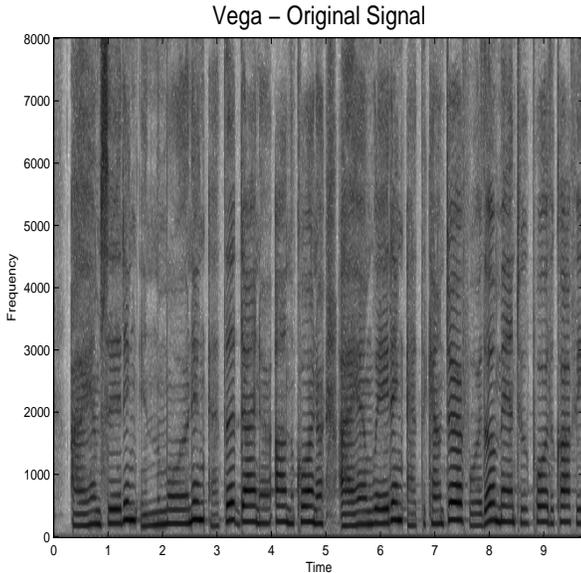


Fig. 6. Original signal spectrogram.

## V. SUMMARY AND CONCLUSIONS

In this paper we presented a method for analysis of audio signals that is based on comparison between two types of parametric spectra - AR spectrum and MVDR spectrum. Summarizing the advantages of our method, we can say that:

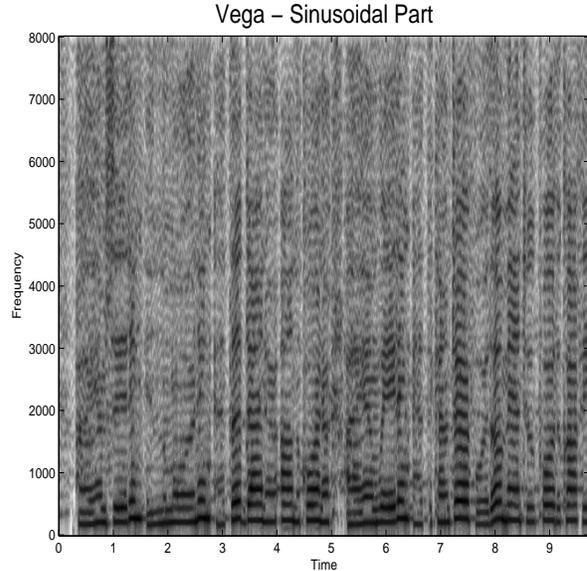


Fig. 7. Sinusoidal component spectrogram.

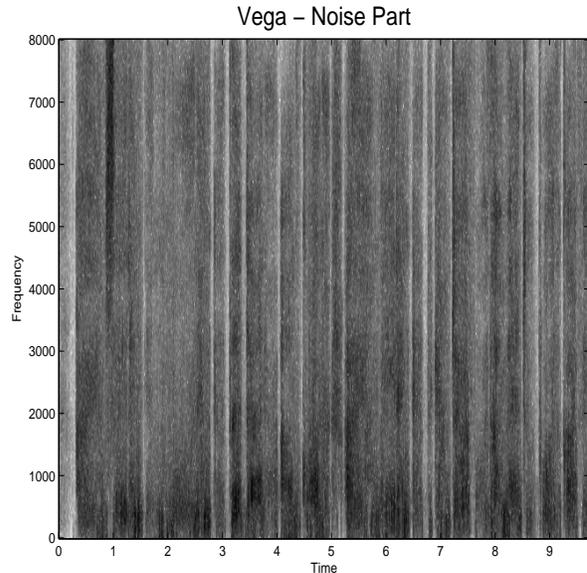


Fig. 8. Noise component spectrogram.

- Parametric spectral methods better utilize signal statistics and are more robust in comparison to Fourier methods that use single snapshot for every frame<sup>2</sup>.
- The AR spectral method allows precise (high resolution) detection of sinusoidal component frequencies without interpolation or zero padding.
- The MVDR method gives optimal estimates of the sinusoidal magnitudes.

<sup>2</sup>It should be noted that averaging of Fourier estimates is possible using shorter FFTs, which may add to robustness but on the expense of proportionally reducing the spectral resolution.

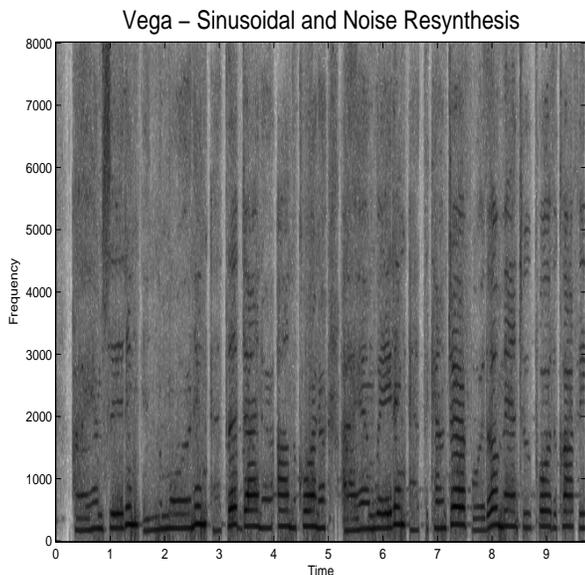


Fig. 9. Combined sinusoidal and noise spectrogram.

- Overfitting of AR model allows modeling of signals that are mixture of sinusoids and noise.
- Bandwidth of modulation of sinusoidal components can be estimated from pole radius.
- Relative weights of the sinusoidal and wide-band noise components are derived from the ratio of spectral powers of the AR and MVDR models.

#### ACKNOWLEDGMENT

I would like to thank Ted Apel for his comments and proofs of the paper.

#### VI. REFERENCES

- [1] R. J. McAulay and Thomas F. Quatieri, "Speech analysis/ synthesis based on a sinusoidal representation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 744–754, August 1986.
- [2] J. W. Beauchamp, "Unix workstation software for analysis, graphics, modification, and synthesis of musical sounds," *Audio Engineering Society, Preprint no. 3479*, 1993.
- [3] D.W. Griffith and J.S. Lim, "Multiband-excitation vocoder," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 2, pp. 236–243, 1988.
- [4] J. Settle and C. Lippe, "Real-time musical applications using the fft-based resynthesis," *Proceedings of the International Computer Music Conference*, 1994.
- [5] X. Serra and J. O. Smith, "Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition," *Computer Music Journal*, vol. 14, no. 4, pp. 12–24, 1990.
- [6] K. Fitz and L. Haken, "Sinusoidal modeling and manipulation using lemur," *Computer Music Journal*, vol. 20, no. 4, pp. 44–59, 1996.
- [7] G. Peeters and X. Rodet, "Sinola: A new analysis/synthesis method using spectrum peaks, harmonic distortion, phase and reassigned spectrum," in *Proc. ICMC*, 1999, pp. 153–156.
- [8] Y. Stylianou, J. Laroche and E. Moulines, "High-quality speech modification based on a harmonic+noise model," in *Proc. EUROSPEECH*, 1995, pp. 451–454.
- [9] H. Purnhagen and N. Meine, "Hiln - the mpeg-4 parametric audio coding tools," in *IEEE International Symposium on Circuits and Systems*, May 2000.
- [10] B. Porat, *Digital Processing of Random Signals*, Prentice-Hall, 1994.
- [11] J. Makhoul, "Linear prediction—a tutorial review," *Proc. IEEE*, vol. 63, no. 2, pp. 561–580, April 1975.
- [12] J. Capon, "High-resolution frequency - wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.
- [13] M.N. Murthi and B.D. Rao, "All-pole modeling of speech based on the minimum variance distortionless response spectrum," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 3, pp. 221–239, May 2000.
- [14] A. Ben-Shalom and S. Dubnov, "Optimal filtering of an instrument sound in a mixed recording using harmonic model and score alignment," in *Proc. ICMC*, Miami, Nov. 2004.
- [15] J. Tabrikian, E. Fisher and S. Dubnov, "Generalized likelihood ratio test for voiced-unvoiced decision in noisy speech using the harmonic model," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 2, pp. 502–510, March 2005.
- [16] B.R. Musicus, "Fast mlm power spectrum estimation from uniformly spaced correlations," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, no. 5, pp. 133–135, Oct. 1985.
- [17] J.P. Burg, "The relationship between maximum entropy spectra and maximum likelihood spectra," *Geophysics*, vol. 37, pp. 375–376, 1972.
- [18] P. Hanna and M. Desainte-Catherine, "Detection of sinusoidal components in sounds using statistical analysis of intensity fluctuations," *Proceedings of the International Computer Music Conference*, 2002.