ANALYSIS OF TRANSIENT MUSICAL SOUNDS BY AUTO-REGRESSIVE MODELING

Florian Keiler[1], *Can Karadogan*[1], *Udo Zölzer*[1], *and Albrecht Schneider*[2]

[1]Dept. of Signal Processing and Communications, Univ. of Federal Armed Forces Hamburg, Germany
[2]Dept. of Systematic Musicology, University of Hamburg, Germany
{florian.keiler,karadog,udo.zoelzer}@unibw-hamburg.de
aschneid@uni-hamburg.de

ABSTRACT

This paper gives an example of an auto-regressive (AR) spectral analysis on transient musical sounds. The attack part of many musical sounds is mostly too short to be analysed by a short-time Fourier analysis, whereas this short period of time is long enough for several AR-methods. The AR-spectra obtained from short segments of signals with attack transients have a sufficiently high frequency resolution. These spectra contain more information about the evolution of a sound than a fast Fourier transform made over a small amount of samples.

1. INTRODUCTION

It is well known that in sounds of many musical instruments attack transients are essential as far as discrimination and identification of various instruments by listeners is concerned. Transients are of interest also from an acoustical point of view since they reflect the vibrational behavior of particular instruments. In many cases, normal modes of vibration are unstable as regards vibrational frequency and/or amplitude during the transient portion of a sound radiated from an instrument. Reasons for such behavior are manifold. For example, in idiophones such as bells, gongs, xylophones etc. which are struck with clappers, mallets etc. energy is transferred by means of an impact causing a sudden deformation of the shape of shells, plates, and bars, respectively. Due to strong impacts, in solids such as bars, plates and shells many modes are elicited which belong to distinct types of vibration (longitudinal, transversal, axial, torsional) and are often found to interact in a complex pattern. In wind instruments, plucked and bowed strings, transients in general occur before a stable regime of standing waves due to resonance is established.

Transients typically include noisy components which can be attributed to the interaction of solids (clapper: bell, mallet: gong), the contact between plectrum and string or hammer and string, or the hiss which comes from blowing and the air flow through valves and tubes in wind instruments.

Attack transients have been investigated in sounds recorded from many instruments (e.g., harpsichords, pianos, guitars, organ flute and reed pipes) with the aid of various techniques. Besides tools widely used such as short-time FFT and wavelet analysis, also autoregressive spectrum analysis (AR) has been employed in the study of piano attack transients [1] as well as in non-western idiophone sounds [2, 3]. The reason to make use of AR techniques simply is that the attack portion in many sounds spans only a few milliseconds (in general, $3\text{ ms} < t < 80\text{ ms}$) whereby it is often difficult to obtain spectra with sufficient resolution from short-time FFT. The aim of a sonological analysis of attack transients is to yield high resolution frequency spectra from rather short segments of the sound signal, and to possibly trace shifts in frequency which are typical of partials during attack.

2. AUTO-REGRESSIVE (AR) MODELLING

An Auto-Regressive (AR) Model can be described by the transfer function

$$H(z) = \frac{1}{1 - \sum_{k=1}^{p} a_k z^{-k}}. \tag{1}$$

The magnitude $|H(e^{j\Omega})|$ models the spectrum of the analyzed signal. This model is commonly used in linear predictive coding (LPC) [4]. AR-models are also called *all-pole models* since the transfer function has only poles (in addition to zeros at $z = 0$), thus, it is a pure IIR-model. In Equation (1) p denotes the model order. The coefficients a_k ($k = 1, \ldots, p$) are calculated from a block of N samples of the input signal.

There exist several methods to compute the coefficients a_k as: Burg algorithm, autocorrelation method, modified covariance (modcov) method [5]. In this paper we focus on the Burg algorithm and the modified covariance method. Both methods are based on the minimization of the sum of forward and backward prediction error energies. The modified covariance uses a direct (transversal) filter structure and results in a linear equation system (normal equation). The Burg method is based on a lattice filter structure and calculates recursively the filter coefficients of successive orders. The modified covariance method may compute unstable synthesis filters $H(z)$ while the Burg method is guaranteed to compute stable filters. This fact is not of interest for the considered task since we are only interested in the spectrum.

Since the prediction order is one of the most important parameters, the restrictions of both methods should be mentioned. The maximum prediction order p depending on the block length N for both methods are given in Equations (2) and (3).

$$p_{\text{max,modcov}} = \frac{2}{3}N, \tag{2}$$

$$p_{\text{max,Burg}} = N - 1. \tag{3}$$

For stationary signals with fundamental frequency f_1 sampled at f_S normally a block length $N = f_S/f_1$ corresponding to the period length is sufficient. The restrictions for the maximum prediction order sometimes require a higher block length in order to get a satisfying AR spectrum.

3. DERIVATIVE ALGORITHM

In order to have a comparison with the FFT approach, the sounds are also analyzed with the derivative algorithm [6, 7].This method computes the FFT of the input signal and a second FFT from the derivative of the input signal and results in improved values of frequency and amplitude of the peaks in the magnitude FFT spectrum. It provides a better detection of sinusoidal components.

The results in [6] showed that the derivative FFT requires a certain frequency difference

$$\Delta f = 2\frac{f_S}{N_{\text{FFT}}} \tag{4}$$

between two frequency components to be able to detect them separately[1], where f_S denotes the sampling frequency and N_{FFT} is the FFT block length. Table 1 lists the frequency detection thresholds for different FFT block lengths at $f_S = 48$ kHz.

N_{FFT} (in samples)	1024	2048	4096
Δf in Hz	93.7	46.8	23.4

Table 1: Frequency detection thresholds of the derivative algorithm for different FFT block lengths at $f_S = 48$ kHz.

4. ANALYSIS

In this section the AR methods are considered for some synthetic signals before we concentrate on the analysis of different natural sounds. All used sounds (synthetic and natural) are represented with a 16 bit resolution at the sampling frequency $f_S = 48$ kHz.

4.1. Synthetic Signals

In order to understand the behavior of the AR models, a comparison between the Burg method and the modified covariance method with synthetic test signals is made.

4.1.1. Stable Sinusoidal

The first example is a pure sinusoid with a single frequency component at $f_0 = 1000$ Hz. The period of the time-discrete signal is 48 samples. An analysis with both AR methods is made with the block length $N = 50$ and the prediction order $p = 10$. Since one resonance frequency (formant) of the AR model spectral envelope requires two complex conjugate poles, $p = 2$ would be sufficient for the analysis of the single sinusoidal. But the frequency spectra for $p = 10$ are overestimated. Figure 1 shows the resulting AR spectra and pole planes with the Burg (top) and modcov (bottom) methods zoomed to the interesting region. In the pole planes the dotted line corresponds to the unit circle while the straight line points to the position of the used sine frequency. The Burg method results into two poles very close to 1000 Hz. The AR spectrum has therefore two peaks around 1000 Hz where only a single peak should be present. This effect is called *spectral line splitting* [5]. This effect cannot be seen at the frequency spectrum of the modified covariance method.

[1]Although the frequency resolution is f_S/N_{FFT}, two local maxima in the FFT magnitude must have the distance Δf.

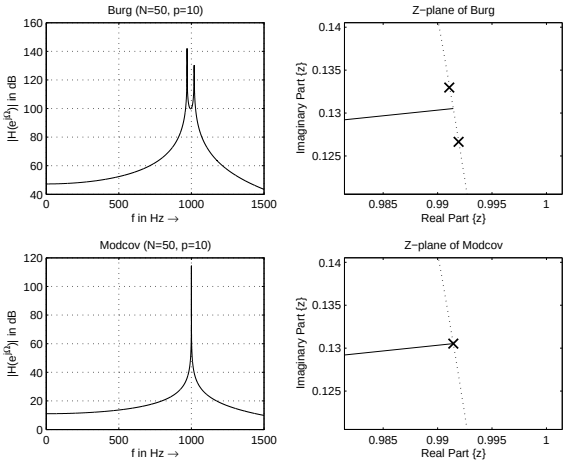

Figure 1: *A zoom-in into the spectra and pole planes of the Burg and the modcov methods with $N = 50$ and $p = 10$.*

4.1.2. Vibrato Sinusoidal

Another effect can be seen at the analysis of a synthetic vibrato signal with a center frequency of $f_0 = 500$ Hz, a frequency depth of $\Delta f = 50$ Hz and a vibrato frequency of 10 Hz. The multisegment spectra[2] of the Burg method can be seen in Figure 2(a). These spectra show a frequency shift over the vibrato character which is caused by the initial phase of the signal [5]. The detected peaks are shifted in terms of a sinusoidal function.

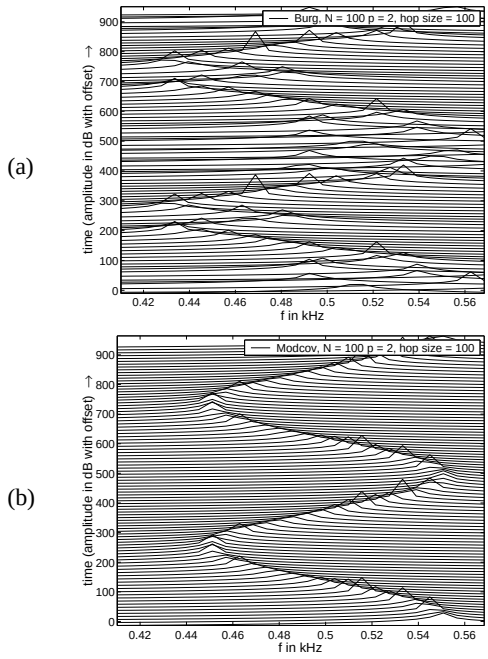

Figure 2: *A zoom-in into the multisegment spectra with $N = 100$ and $p = 2$ for the Burg (a) and modcov (b) methods.*

[2]This is a diagram type showing subsequent frequency spectra plotted after another.

This effect cannot be seen at the multisegment spectra of the modified covariance method depicted in Figure 2(b). The method can track the frequency of the vibrato signal properly. Therefore, the modcov method is more suitable for analyzing natural sounds.

4.2. Natural Sounds

4.2.1. A Carillon Bell From Ghent

A bell sound recorded from one of the finest extant historical carillon bells (cast by Pieter Hemony in Ghent around 1660) will be analyzed first. The bell is the second largest in the Ghent carillon comprising 52 bells. The sound (48 kHz/16 bit) was recorded to DAT with a condenser microphone put at a distance of appr. 100 cm to the bell's rim. Figure 3 shows the first 50 ms (2400 samples) of the time signal. The FFT spectrum calculated for the first 8192 samples from onset yields a pattern of partials which is typical for minor-third bells. The frequencies of the strongest partials up to 1.2 kHz which were obtained from estimation of spectral peaks by means of parabolic interpolation[3] are listed in Table 2.

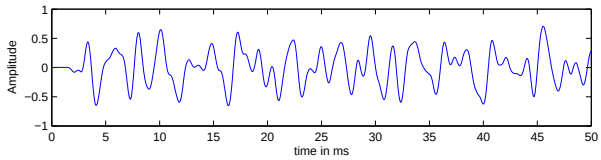

Figure 3: *The first* 50 *ms of the Ghent Bell time signal.*

Partial no. i	Freq. f_i in Hz	Ampl. in dB	Ratio f_i/f_1
1	106.89	-32.5	1.0000
2	214.21	-31.1	2.0040
3	255.60	-22.6	2.3912
4	426.76	-22.6	3.9925
5	481.74	-46.7	4.5068
6	552.51	-32.4	5.1689
7a	597.51	-43.7	5.5899
7b	609.07	-47.3	5.6981
8	639.61	-25.7	5.9838
9	698.57	-45.8	6.5354
10	732.71	-57.3	6.8547
11	791.03	-51.8	7.4004
12	884.29	-27.7	8.2728
13	908.35	-54.8	8.4979
14	927.27	-49.3	8.6749
15a	971.90	-55.0	9.0924
15b	980.45	-57.1	9.1725
16	994.22	-59.1	9.3012
17	1010.36	-50.8	9.4523
18	1035.36	-54.6	9.6861
19	1060.63	-65.4	9.9225
20	1116.45	-64.4	10.4447
21	1152.29	-43.3	10.7800

Table 2: Significant spectral peaks (0–1.2 kHz) of Ghent bell no. 2 calculated by FFT with $N_{\text{FFT}} = 8192$ and parabolic interpolation.

[3]Calculated with the "Spectro" program from CCRMA, Stanford. Similar results are obtained with the derivative algorithm.

Frequency ratios between some of the partials are harmonic (or nearly so). However, The spectrum contains also inharmonic components the number of which increases within the third and fourth octaves above the fundamental. Because of spectral inharmonicity, the time function of the sound is complex. Frequency pairs 7a/b and 15a/b denote degenerate pairs of eigenmodes due to small deviations from the axisymmetrical mass distribution of the bell [8].

The vibrational behavior of a bell and the spectral composition of bell sounds depends largely on the impact force which is transmitted to the bell by means of a clapper (contact time 1 ms $\leq t <$ 5 ms). The temporal evolution of bell sound spectra can be studied using STFT technique, however, with a frequency resolution defined by f_S/N_{FFT}, the time window in many cases will need at least a block length of 2048 or even 4096 samples. As an alternative, AR-models offer a better temporal resolution.

To find most of the stronger partials, an AR analysis based on the modcov method with a block length of $N = 1024$ and a prediction order of $p = 510$ is sufficient, see results in Tab. 3. The period length of the sound is $N_T = f_S \cdot 9.36$ ms ≈ 449 samples.

The modcov results show that the spectrum of the bell builds up quite rapidly due to the impact. The time series of $N = 1024$ samples already contains the partials also found in a much longer FFT window (Table 2). Deviations in frequency between the FFT analysis with $N_{\text{FFT}} = 8192$ listed in Table 2 and the modcov analysis listed in Table 3 may in part be attributed to slight glides in partial frequencies often found at the onset of idiophone sounds.

Partial no. i	Freq. $\tilde{f}_i$ in Hz	Ratio $\tilde{f}_i/\tilde{f}_1$	Ratio $\tilde{f}_i/f_1$
1	109.62	1.0000	1.0255
2	214.84	1.9599	2.0099
3	254.62	2.3228	2.3821
4	415.12	3.7869	3.8836
5	451.72	4.1208	4.2260
6	564.29	5.1477	5.2792
7	648.39	5.9149	6.0660
8	821.76	7.4964	7.6879
9	882.71	8.0525	8.2581
10	966.70	8.8186	9.0439
11	1116.54	10.1856	10.4457

Table 3: Spectral peaks (0–1.2 kHz) of Ghent carillon bell no. 2 from modcov AR spectrum with $N = 1024$ and $p = 510$.

4.2.2. Plucked String of a Harpsichord

In harpsichords, depressing a key leads to raising the respective jack which carries a small plectrum. This plucks the string which is first raised, and then slips from the tip of the plectrum. Besides transversal motion of the string, longitudinal modes can be expected if the excitation is strong enough. Also, some torsional motion may occur due to the string slipping from the plectrum which itself is bending downwards.

As an example, a sound from a Kirkman harpsichord built in 1766 will be analyzed. We played the A1 string tuned to appr. $f_1 =$ 51 Hz leading to the period $T \approx 19.6$ ms (941 samples).

In the time function of sounds recorded from single notes played on a harpsichord, typically three or four segments can be distinguished (see Figure 4): segment (A) covers the time needed to put the plectrum against the string and to raise the string before

it slips from the plectrum; (B) is the time immediately after the string has been released from the plectrum; during (C), the first quasi-period of length $T \approx 1/f_1$ reflecting the fundamental frequency the string is tuned to emerges; finally, (D) begins where periodic motion of the string is established.

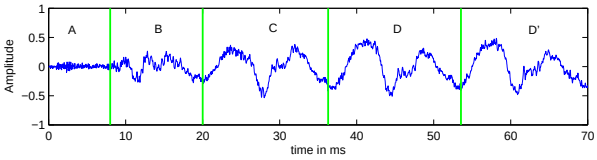

Figure 4: *Time signal of the Kirkman harpsichord with segments A–D.*

The sound during (A) is noisy since it is produced by mechanical interaction of the key, jack, plectrum, and string. In (B), the sound is transient because periodic motion of the string is not yet established. In (C), quasi-periodic motion of the string produces likewise a time function which already resembles a "true" period which will be found later in D with only small variation in period length. Whereas (A) shows a noise-like structure, the spectral composition of, in particular, (B) is difficult to investigate due to the transient, non-periodic nature of this sound segment which in many cases spans only a few microseconds (depending mainly on the fundamental frequency of the string). From (C) on, periodicity appears in the signal which in (D) becomes more stable.

To investigate (A), (B) as well as the beginning of (C), normal STFT will not work very well since the sound segments are typically too short to allow a FFT window of length $N = 2048$ (or longer) which will provide sufficient frequency resolution. We therefore have applied the AR modcov method to study the temporal and spectral evolution of harpsichord sounds. The analysis results are shown in Figure 5. The shown FFT spectra were computed with a block length $N_{\text{FFT}} = 2048$. The middle of the FFT and AR blocks were identical.

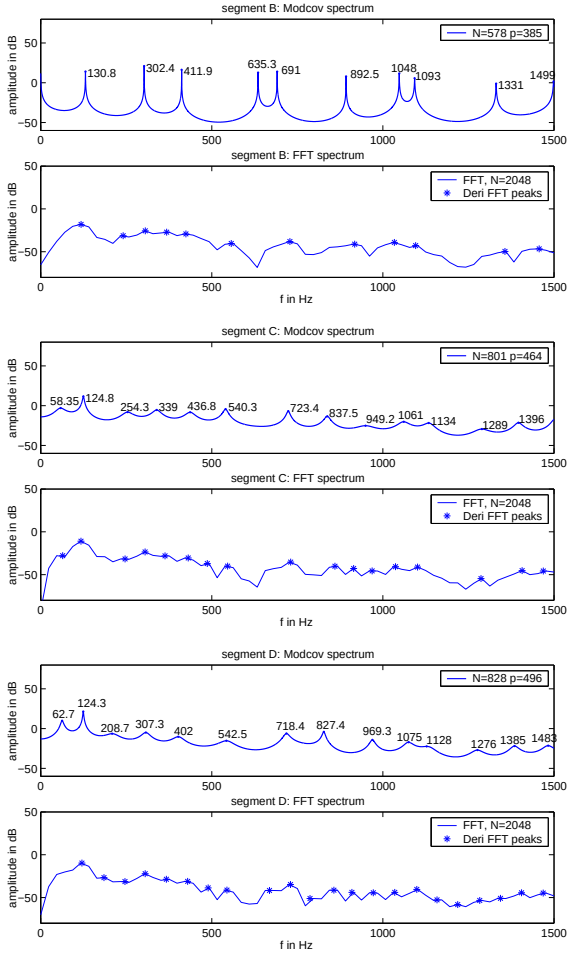

Figure 5: *The analysis of the Kirkman harpsichord of segments B, C, D with modcov and derivative FFT, respectively. The numbers in the AR spectra denote the maxima positions as frequency in Hz.*

5. CONCLUSION

Depending on the prediction order and the block length, an AR analysis can locate the peaks of a short-time signal block more accurately than an FFT analysis of the same short length. Using shorter block lengths, a better time resolution can be obtained and variations of frequency components can be detected. This major advantage of the AR approach over the FFT can be taken when the behavior of various AR methods are known and the appropriate prediction order and block length for the sound signal can be estimated.

With synthetic signals the Burg algorithm shows some inaccuracies which do not occur with the modified covariance (modcov) method. Since there is always a possibility that the AR-spectrum may be overestimated, one should also consider for comparison the short-time FFT with different block lengths and the results of the derivative algorithm. These results can be combined with the AR results for analyzing natural sounds.

6. REFERENCES

[1] W. Bachmann, H. Bücker, and B. Kohl, "Feinstrukturanalyse des Einschwingens eines Pianoklanges," *Acustica*, vol. 68, pp. 123–130, 1989.

[2] A. Schneider, *Tonhöhe - Skala - Klang*, Orpheus, Bonn, 1997.

[3] A. Schneider, "Autoregressive spectral analysis and wavelet gammatone filtering of idiophone sounds: implications for pitch perception," in *Music and Signs*, I. Zannos, Ed., pp. 99–116. Asco Publ., Bratislava, 1999.

[4] John Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, Apr. 1975.

[5] S. Lawrence Marple, Jr., *Digital Spectral Analysis with Applications*, Prentice Hall, Englewood Cliffs, New Jersey, 1987.

[6] F. Keiler and S. Marchand, "Survey On Extraction of Sinusoids in Stationary Sounds," in *Proc. DAFx-02 Digital Audio Effects Conference*, Hamburg, Sept. 2002, pp. 51–58.

[7] Myriam Desainte-Catherine and Sylvain Marchand, "High Precision Fourier Analysis of Sounds Using Signal Derivatives," *JAES*, vol. 48, no. 7/8, pp. 654–667, July/August 2000.

[8] A. Schneider and M. Leman, "Sonological and psychoacoustic characteristics of carillon bells," in *Proc. of the sixteenth meeting of the Foundations of Music [FWO] Research Soc., The Quality of bells* , IPEM, Ghent University, Ed., Brugge, Belgium, Sept. 2002.