ARMA PROCESSES IN SUBBANDS WITH APPLICATION TO AUDIO RESTORATION

Luiz W. P. Biscainho, Paulo S. R. Diniz

Universidade Federal do Rio de Janeiro
LPS - DEL-EE & PEE-COPPE
Caixa Postal 68564
21945-970 Rio de Janeiro, RJ, Brazil
{wagner,diniz}@lps.ufrj.br

Paulo A. A. Esquef

Helsinki University of Technology
Lab. of Acoustics and Audio Signal Proc.
P.O. Box 3000
FIN-02015 Espoo, Finland
esquef@acoustics.hut.fi

ABSTRACT

It was recently shown [1] that the sub-band signals resulting from the analysis of an ARMA process by a decimating filter bank can be modeled by individual lower-rate ARMA processes, by using the power spectral density concept. This paper discusses the validity of this modeling and the mapping of zeros and poles of the corresponding generator filters along the sub-bands. Based on these results, after some considerations related to the employment of sub-band systems in audio restoration, a sub-band version of a model-based technique for suppression of impulsive noise from audio recordings is proposed. Simulation results are presented and validated by subjective as well as objective means.

1. INTRODUCTION

Multi-rate digital processing [2] has been extensively applied in several areas, including speech and audio processing. The class of ARMA (autoregressive moving average) random processes [3] is commonly employed as a model for speech and audio signals.

Enhancement or restoration of audio recordings [4] deals with problems broadly classified as localized and global disturbances, of which the most common are clicks (impulsive noise) and hiss (broadband noise), respectively. Click suppression methods often uses AR (autoregressive) models [5, 4] successfully, whereas hiss reduction methods are frequently based, directly or indirectly, on some multi-rate approach [5, 6].

It was recently derived an equivalent lower-rate ARMA model for each sub-band signal resulting from the analysis of an ARMA process by a decimating filter bank [1], which resulted in a potential tool for audio restoration through model-based multi-rate processing. The present work briefly reviews this form of modeling and discusses its validity. Possible model-order reduction is investigated by mapping the generator-filter zeros and poles along the sub-bands.

Justified by the existence of equivalent sub-band ARMA models, we implement a multi-rate system for suppression of impulsive noise from audio recordings in which each sub-band signal is processed by an improved version [7] of the model-based technique first proposed in [5]. Through simulated examples, the method is evaluated in terms of complexity and attained final quality, the latter assessed by subjective as well as objective means based on the perceptual measure described in [8].

The paper is organized as follows. Section 2 reviews the model for an ARMA process in sub-bands and discusses related issues.

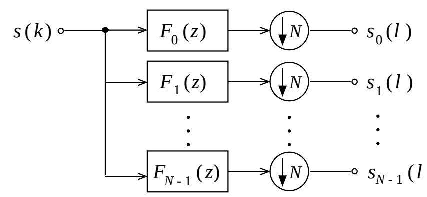

Figure 1: ARMA Process analyzed by an N-band decimating filter bank.

After considerations on the application of sub-band systems to audio restoration, Section 3 proposes and evaluates a sub-band algorithm for click suppression. Conclusions are drawn in Section 4.

2. ARMA PROCESS IN SUB-BANDS

2.1. Model: Lower-Rate ARMA Processes

Consider the N-band multi-rate system shown in Fig. 1, which analyzes the ARMA process defined by

$$s(k) = \sum_{i=1}^{I} a(i)s(k-i) + \sum_{j=0}^{J} b(j)e(k-j),$$

where $e(k)$ is white noise with variance σ_e^2.

For a given sub-band n, we can group the process generating filter with the analysis filter (Fig. 2a) and obtain the corresponding polyphase representation (Fig. 2b). In [1], it was shown that, since $e_p(l)$, $p = 0, ..., N-1$, are mutually uncorrelated white-noise processes with the same variance as $e(k)$, $s_n(l)$ can be modeled by an equivalent ARMA process $\tilde{s}_n(l)$ in terms of its power spectral density. This process is obtained applying a white-noise process $e_{d_n}(l)$ with variance σ_e^2 as input to the minimum-phase filter $\hat{F}_{d_n}(z)$ (Fig. 2c) that results from spectral factorization of

$$\begin{aligned} \mathcal{S}_{s_n s_n}(z) &= \sum_{p=0}^{N-1} \hat{F}_{np}(z)\hat{F}_{np}(z^{-1})\sigma_e^2 \triangleq \\ &\triangleq \hat{F}_{d_n}(z)\hat{F}_{d_n}(z^{-1})\sigma_e^2. \end{aligned} \tag{1}$$

Now, the equivalent model for the multi-rate system would consist of N similar subsystems in parallel. Completing the model, the cross-power spectral density between two distinct sub-bands

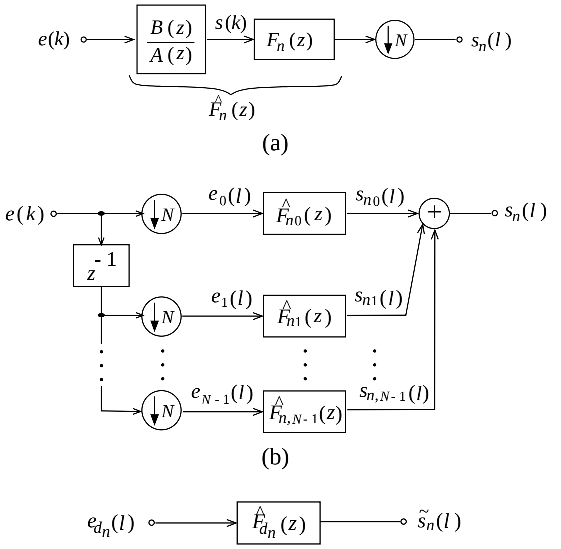

Figure 2: Generation of $s_n(l)$: from the multi-rate representation to a low-rate model.

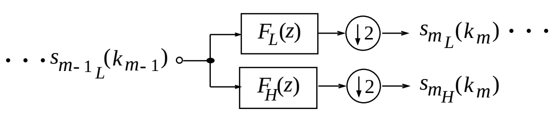

Figure 3: The m-th stage of an octave filter bank.

n_1 and n_2, from which their cross-correlation can be obtained, is given [1] by

$$\mathcal{S}_{s_{n_1} s_{n_2}}(z) = \left[\sum_{p=0}^{N-1} \hat{F}_{n_1 p}(z) \hat{F}_{n_2 p}(z^{-1}) \right] \sigma_e^2.$$

2.2. Discussion

A special case related to audio processing is an AR process being split in octaves by a binary tree-structured filter bank formed by successive stages similar to that shown in Fig. 3. The previous results can be easily adapted to this environment, since AR models are a particular case of ARMA models and the octave sub-band models can be computed recursively using equation (1) for each decimating stage [1]. In this Section, for the sake of simplicity, we refer only to this case and assume $F_L(z)$ and $F_H(z)$ are FIR filters. The generating filters associated to the equivalent processes $\tilde{s}_{m_L}(k_m)$ and $\tilde{s}_{m_H}(k_m)$ will be called respectively $\hat{F}_{d_{m_L}}(z)$ and $\hat{F}_{d_{m_H}}(z)$.

2.2.1. The Model Validity

Assuming that ARMA models can adequately represent the sub-band signals, do the model filters computed by equation (1) match the frequency responses of the filters obtained by direct estimation of ARMA parameters from those signals? Simulations demonstrate that the two frequency responses are virtually indistinguishable, except for limitations of the parameter estimator itself, e.g. when performing estimation from insufficient number of samples. Also, replacement of ARMA models by AR models with the same denominator order has negligible effect, in general.

Simulations also confirm the sub-band ARMA models adequacy, since each sub-band signal, when applied as input to its inverse model filter, is essentially whitened.

As an example, we analyzed by an octave Haar filter bank an AR process with poles obtained by combination of magnitudes 0.5 and 0.99 with phases 0, $\pm\frac{\pi}{18}$, $\pm\frac{\pi}{3}$, $\pm\frac{\pi}{2}$, $\pm 2\frac{\pi}{3}$ and $17\frac{\pi}{18}$. Fig. 4 compares the magnitude and phase responses of $\hat{F}_{d_{2_L}}(z)$ associated to the ARMA model computed recursively by equation (1), and the ARMA and AR models with the same denominator order obtained by direct estimation from sub-band signals from a simulation. The signal obtained by inverse filtering $s_{2_L}(k_2)$ through its model-predicted generator filter exhibits the normalized autocorrelation shown in Fig. 5.

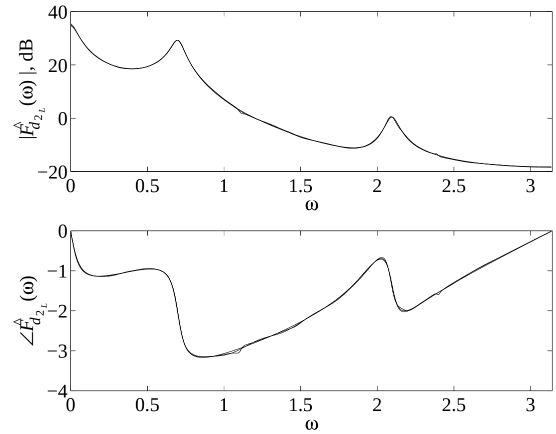

Figure 4: Magnitude and phase responses of $\hat{F}_{d_{2_L}}(z)$: theoretical ARMA model and estimated ARMA and AR models.

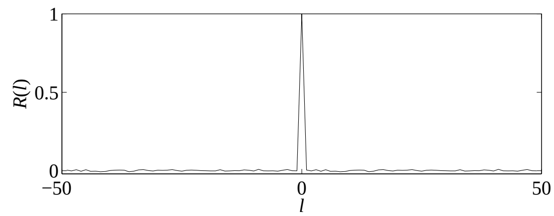

Figure 5: Inverse model-filtering of $s_{2_H}(k_2)$: autocorrelation.

2.2.2. Mapping of Zeros/Poles along the Sub-bands

Suppose that $F_L(z)$ and $F_H(z)$ are ideal half-band low- and high-pass filters, respectively. A cosine wave with frequency ω applied as input to the octave filter bank would appear in $s_{m_L}(k_m)$ with its frequency doubled after each stage m, until being separated in the signal $s_{m_\omega H}(k_m)$, after stage

$$m_\omega = \left\lceil \log_2 \frac{\pi}{2\omega} \right\rceil + 1,$$

at this point with frequency $2\pi - 2^{m_\omega}\omega$. Now, a general AR process $s(k)$ with poles z_{p_i} analyzed by a single-stage octave filter bank would result in two ARMA processes, each one with poles $z_{p_i}^2$ (implying that phases are doubled and magnitudes are squared) and the necessary zeros to perfectly separate half power spectrum, with the zeros also squared. The same effect occurs along further stages. Therefore, the effect of each original pole cannot be confined to an only sub-band, except for the limit case of unit magnitude (corresponding to the cosine wave).

For realizable filter banks the pole squaring property—due to decimation-by-2—is preserved. Zero mapping is hard to predict without calculations; however, some speculations can be made.

First, it is possible to deduce that the numerator and denominator orders of $\hat{F}_{d_{m_L}}(z)$ or $\hat{F}_{d_{m_H}}(z)$ are

$$\begin{cases} O_{\text{num}}(m) = \left\lceil \frac{O_{\text{num}}(m-1)+I}{2} \right\rceil, \ \text{m} \geq 2 \\ O_{\text{num}}(1) = \left\lceil \frac{Q+I}{2} \right\rceil \end{cases}$$

and $O_{\text{den}}(m) = I$, respectively, except for possible canceling terms, where I is the order of the AR process under analysis and Q is the order of the FIR filters $F_L(z)$ and $F_H(z)$.

Disregarding decimation for the moment and considering ideal filter banks, the bandwidth of the original AR process $s(k)$ is increasingly narrowed along the analysis stages. At a given stage, the individual effect of any pole of the generator filter of $s(k)$ distributes itself between passband and stop-band of the corresponding equivalent filters; if it is almost entirely contained in a stop-band, its strong rejection can be achieved by its canceling by a zero almost in the same position. Recalling decimation and non-ideal filter banks, we add that after the m-th stage each pole is represented by its m-times squared version and that some aliasing occurs, but the idea remains the same. As a consequence, along successive stages, one expects that the equivalent model generator filters have an increasing number of poles canceled by zeros. Note that, although the filter banks are fixed, the calculations of the numerators of $\hat{F}_{d_{m_L}}(z)$ (or $\hat{F}_{d_{m_H}}(z)$) combine the poles of $s(k)$ with the sub-filters $F_L(z)$ (or $F_H(z)$), in such a way that the zeros can 'chase' the poles.

This fact, associated to the successive reduction of pole magnitudes, suggests the possibility of a gradual reduction of the model orders along the stages without impairing the overall accuracy [9].

As an illustrative example, the octave analysis of a 40th order AR process was performed by a Haar filter bank. Fig. 6 depicts the calculated zero-pole diagram of $\hat{F}_{d_{4_H}}(z)$, showing several poles canceled by zeros. After computing the complete models, any individual zero, pole or zero-pole pair whose associated frequency response deviated less than 10% from unit magnitude and less than $\frac{\pi}{10}$ from linear phase was discarded, since they would have little effect on the overall transfer function. Table 1 shows the corresponding complete and simplified model orders, with the latter greatly reduced. One could further think of replacing these low-order ARMA models by sufficient-order AR models, which are simpler to estimate.

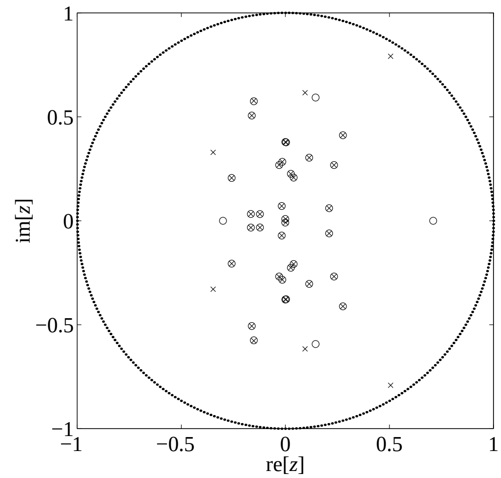

Figure 6: AR process octave analysis: zero-pole diagram of $\hat{F}_{d_{4_H}}(z)$.

Table 1: A 40th order AR process analyzed in octaves: number of zeros and poles of the model generator filters.

	complete		reduced	
filter	# zeros	# poles	# zeros	# poles
$\hat{F}_{d_{1_H}}(z)$	20	40	6	26
$\hat{F}_{d_{1_L}}(z)$	20	40	3	27
$\hat{F}_{d_{2_H}}(z)$	30	40	2	12
$\hat{F}_{d_{2_L}}(z)$	30	40	1	12
$\hat{F}_{d_{3_H}}(z)$	35	40	3	6
$\hat{F}_{d_{3_L}}(z)$	35	40	1	12
$\hat{F}_{d_{4_H}}(z)$	38	40	2	4
$\hat{F}_{d_{4_L}}(z)$	38	40	1	4

3. APPLICATION TO AUDIO RESTORATION

3.1. Advantages of sub-band audio enhancement

The same way modern audio coding and compression apply psychoacoustics principles [8] to know 'How worse can the audio be made, before this can be perceived?', audio enhancement could use them to know 'How better can the audio be made, before this cannot be perceived?' in order to reduce computation and loss of quality by over-processing. A main issue is Masking, the property a sound has to inhibit perception of another one, near to the former in time and/or frequency. It is known that only a narrow frequency band around a tone (called Critical Band) contributes to its masking. A fixed arrangement of critical bands is usually adopted to model sound perception, and leads to a corresponding sub-band configuration.

Simultaneous sub-band hiss reduction and model-based click suppression can be attempted by an adequate multi-rate system.

As disturbances may sometimes be found in preferential frequency bands, one can expect the computational burden to be reduced and over-processing to be avoided by sub-band processing, since more 'attention' will be given to the most corrupted bands.

Other processing savings are possible in the event of a favorable equivalence between complex operations in full band and simpler ones in sub-bands.

3.2. Example: Click Suppression in Sub-bands

A Least-Squares (LS)-based technique for suppression of clicks from audio recordings, first proposed in [5] and reviewed in [4], models each block of N audio samples as a p-th order AR pro-

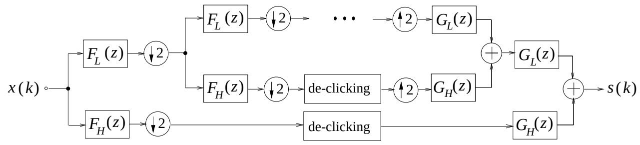

Figure 7: Proposed structure for click suppression in sub-bands. Synchronizing delays were omitted.

cess additively corrupted by impulsive noise. A Detection step is performed first, in which the AR model is LS-estimated, the noisy signal is inverse filtered and a threshold λ, proportional to an estimate of the excitation standard-deviation by a factor K, is computed; all signal samples whose corresponding excitation magnitudes exceed λ are considered as corrupted. In a Correction step, these ones are replaced by LS-interpolated values. In order to improve detection performance, some modifications have been added to this technique, such as using an auxiliary threshold and merging contiguous clicks [7].

Driven by some ideas addressed in Section 3.1 and taking in consideration the possibility of modeling individual sub-band signals by ARMA/AR processes, we propose here a multi-rate system for click suppression which uses the improved De-clicking method described above in each sub-band, as depicted in Fig. 7. In fact, the proposed detection/suppression does not explore mutual correlations between distinct sub-bands; however, it is expected that they are preserved, if one considers the small percentage ($< 10\%$) of samples to be corrected in a typical De-clicking procedure.

In this context, several factors can influence global computational complexity of the system: sub-band block sizes, limited by the maximum stationarity periods which can be assumed; possibility of ARMA model order reduction or adoption of AR model in each sub-band; and option of employing simplified processing (or none, at all) in non critical sub-bands.

Two practical examples are examined in detail. In each one, a high-quality uncorrupted audio recording (to be used as a reference signal in the objective quality measurement) was additively corrupted with real impulsive noise; restoration was performed through the full-band algorithm with $N = 1024$, $p = 40$ and $K = 5$, as well as using our multi-rate system with 4 octave stages of 15th order QMF filter banks, i.e. 5 sub-bands, each one with a 40th order AR model, $N = 512$ and $K = 5$. Only $s_{1_H}(k_1)$, $s_{2_H}(k_2)$ and $s_{3_H}(k_3)$ (where clicks are more evidenced) were processed. Case 1: signal S1 with 0.5% of its samples corrupted; both enhanced versions resulted excellent. Case 2: signal S2 with 5% of its samples corrupted; both enhanced versions resulted much superior to the corrupted signal, but the full-band system performed better. Sub-band processing was between 2 and 5 times faster than full-band. An additional objective evaluation is provided by the PAQM (Perceptual Audio Quality Measure) proposed in [8], which compares a processed and the original version of an audio signal. Equal signals give PAQM=0; in our case, comparing enhancement results with the uncorrupted signal, small values indicate successful restoration. Table 2 shows measurements for corrupted, full-band and multi-band restored versions of S1 and S2, which confirm human judgement.

Table 2: PAQM for test signals S1 and S2.

	S1	S2
corrupted	0.243	1.191
full-band restored	0.027	0.060
sub-band-restored	0.023	0.200

4. CONCLUSIONS

A model for ARMA processes split in sub-bands, consisting of individual lower-rate ARMA processes, was shown to be a valid alternative for practical applications. Zero-pole mapping of sub-band model generator filters indicated the possibility of model order reduction. As an application of this sub-band modeling, a multi-rate De-clicking system was proposed. Practical simulations indicated that the multi-rate solution can achieve performance comparable to that of the full-band algorithm, at reduced computational cost. As a natural sequence of this work, a refined version of the sub-band de-clicking algorithm is being developed which takes in consideration the cross-dependence between sub-bands at the same rate.

5. REFERENCES

[1] L. W. P. Biscainho, P. S. R. Diniz, and P. A. A. Esquef, “A model for an arma process split in sub-bands,” in *Proc. 2000 IEEE Intern. Symposium on Circuits and Systems*, vol. III, (Geneve, Switzerland), IEEE, May 2000.

[2] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, chapter 4, pp. 100–133. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[3] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, section 10.4, pp. 332–336. New York, NY, USA: McGraw-Hill, 3 ed., 1991.

[4] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration*. London, UK: Springer, 1998.

[5] S. V. Vaseghi, *Algorithms for Restoration of Archived Gramophone Recordings*. Ph.D. dissertation, University of Cambridge, Cambridge, UK, Feb. 1988.

[6] J. Berger, R. R. Coifman, and M. J. Goldberg, “Removing noise from music using local trigonometric bases and wavelet packets,” *J. Audio Eng. Soc.*, vol. 42, pp. 808–818, Oct. 1994.

[7] P. A. A. Esquef, L. W. P. Biscainho, P. S. R. Diniz, and F. P. Freeland, “A double-threshold-based approach to impulsive noise detection in audio signals,” in *Proc. 2000 EUSIPCO - European Signal Processing Conference*, 2000.

[8] J. G. Beerends and J. A. Stemerdink, “A perceptual audio quality measure based on a psychoacoustic sound representation,” *J. Audio Eng. Soc.*, vol. 40, pp. 963–978, Dec. 1992.

[9] L. W. P. Biscainho and P. S. R. Diniz, “On the effects of zero-pole pairs and individual zeros and poles on discrete-time transfer functions,” in *Proc. 2000 IEEE Intern. Symposium on Circuits and Systems*, vol. I, (Geneve, Switzerland), pp. 363–366, IEEE, May 2000.