ON THE CONTROL OF THE PHASE OF RESONANT FILTERS WITH APPLICATIONS TO PERCUSSIVE SOUND MODELLING

Mathieu Lagrange, Nathan Whetsell, Philippe Depalle

Music Technology Area,
CIRMMT, McGill University,
Montréal, QC, Canada
mathieu.lagrange@mcgill.ca

ABSTRACT

Source-filter models are widely used in numerous audio processing fields, from speech processing to percussive/contact sound synthesis. The design of filters for these models—be it from scratch or from spectral analysis—usually involves tuning frequency and damping parameters and/or providing an all-pole model of the resonant part of the filter. In this context, and for the modelling of percussive (non-sustained) sounds, a source signal can be estimated from a filtered sound through a time-domain deconvolution process. The result can be plagued with artifacts when resonances exhibit very low bandwidth and lie very close in frequency. We propose in this paper a method that noticeably reduces the artifacts of the deconvolution process through an inter-resonance phase synchronization. Results show that the proposed method is able to design filters inducing fewer artifacts at the expense of a higher dynamic range.

1. INTRODUCTION

The source-filter model is widely used in numerous audio processing fields, from speech processing to percussive/contact sound synthesis. Consequently, estimation of the filter parameters has received much attention in the literature [1]. These filters usually model the body resonances of the sound generation system. They are usually expressed in terms of central frequency and bandwidth (or damping parameter).

The initial phase of the filter, though neglected in many synthesis scenarios, also influences the synthesis quality [2]. As shown in [3] and detailed in the remaining of the paper, controlling the initial phase becomes very important to achieve a high quality estimation of the excitation signal from an actual recorded sound.

Indeed, this estimation can be performed by deconvolving a recorded signal using the inverse of a model filter. Because deconvolution is equivalent to multiplying the recorded signal's spectrum by the inverse of the frequency response of the filter, it is critical that we control the valleys of the filter's magnitude response. Without explicit control of the initial phase, antiresonances in the filter's magnitude response could generate undesirable resonances in the estimated excitation signal.

This issue has been previously identified and partially addressed by Laroche and Meillier in [3]. Their study focused on modelling the piano, the resonant modes of which are quasi-harmonically spaced in the spectrum. Even though they use a cosine section structure for the model filter, the authors notice in [4] that the estimated excitation signal still contains some pitched information. Further analysis shows that this pitch information does not correspond to any filter resonances. Rather, the pitch information is actually due to artificially generated resonances resulting from the inversion of the antiresonances. Since Laroche and Meillier consider harmonic sounds, frequency locations of the antiresonances of the filter are accordingly harmonically spaced but shifted by half of the fundamental frequency, resulting in a distorted pitch perception. In order to tackle this issue, they propose in [3] a method specifically tailored to the piano: joint estimation of the excitation signal from recordings of several chords hit by the same type of hammer at the same velocity.

When dealing with contact sounds such as a hammer hitting a plate, one has to cope with filters of even more complex structure [5]. The geometry of the plate will generate a set of inharmonic resonances whose frequencies may be arbitrarily close. In this case, the antiresonances are accordingly even more pronounced.

We propose in this paper to explicitly adjust the initial phases of each mode of the resonant filter in order to smooth the shape of the valleys of its magnitude response. As detailed in Section 3, we consider the closed-form expression of a simple filter structure consisting of two complex poles, derived in Section 4, to find the phase difference between neighbouring poles. By assuming that the influence of the non-adjacent poles is negligible, we will consider an incremental algorithm to adjust the phases of the modes.

The performance of our proposed method of filter design is compared to state-of-the-art filter structures such as those proposed by Laroche and Meillier in Section 5. The comparisons consider several criteria, such as the preservation of the dynamic range of the filter, the preservation of the amplitude ratio between the different resonances, and the reduction of the antiresonances phenomenon. Results show that the proposed method preserves the ratio of modal amplitudes and achieves a much better reduction of the antiresonances at the expense of a higher dynamic range.

2. THE SOURCE-FILTER MODEL

A percussive sound can be conveniently decomposed into two elements: the excitation and the response. To model such a sound, one can consider an additive scheme [6] where a highly damped transient is superimposed over a slowly-decaying sound. The transient part models the interaction of the exciter and the resonator—the attack of the sound. The slowly-decaying sound represents the resonating structure vibrating according to its own modes.

As remarked in [4], the complete separation of excitation and response may be convenient, but it neglects the coherence of the interaction between the excitator and the resonances. Indeed, the role of the exciter is to distribute energy over the different modes of the vibrating structure. The source-filter model is therefore more appropriate as it explicitly models the physical interaction between excitation and response, though it usually neglects the coupling of the source and the filter.

Assuming that the excitation is short compared to the total duration of the resulting sound (which is generally the case for percussive sounds), we can model the response of the sound as a sum of K damped sinusoids

$$h(n) = u(n)\sum_{k=1}^{K} g_k r_k^n \cos(\theta_k n + \phi_k), \quad (1)$$

where g_k, r_k, θ_k, and ϕ_k, are respectively the gain, damping, angular frequency, and initial phase of sinusoid k; and $u(n)$ denotes the unit step function. Taking the Z-transform of $h(n)$, we obtain

$$H(z) = \frac{1}{2}\sum_{k=1}^{K} g_k\left(\frac{e^{j\phi_k}}{1-r_k e^{j\theta_k}z^{-1}} + \frac{e^{-j\phi_k}}{1-r_k e^{-j\theta_k}z^{-1}}\right) \quad (2)$$

$$= \sum_{k=1}^{K}\frac{B_k(z)}{A_k(z)} = \frac{\sum_{k=1}^{K} B_k(z)\prod_{\ell\neq k} A_\ell(z)}{\prod_{k=1}^{K} A_k(z)} = \frac{B(z)}{A(z)}, \quad (3)$$

where

$$B_k(z) = g_k[\cos(\phi_k) - r_k\cos(\theta_k-\phi_k)z^{-1}] \quad (4)$$

$$A_k(z) = 1 - 2r_k\cos(\theta_k)z^{-1} + r_k^2 z^{-2}. \quad (5)$$

The denominator of the filter $H(z)$ models the resonant part of the percussive signal, while initial phases ϕ_k and gains g_k, though dependent on the excitation, may be set arbitrarily (as the initial conditions in physics).

By setting $\phi_k = -\frac{\pi}{2}$ for all k, we obtain a filter consisting of parallel second-order sine sections [3]:

$$H_{\text{sine}}(z) = \frac{1}{2j}\sum_{k=1}^{K} g_k\left(\frac{1}{1-r_k e^{j\theta_k}z^{-1}} - \frac{1}{1-r_k e^{-j\theta_k}z^{-1}}\right). \quad (6)$$

The impulse response of $H_{\text{sine}}(z)$ consists of a sum of K damped sines. By setting $\phi_k = 0$ for all k, we obtain a filter consisting of parallel second-order cosine sections:

$$H_{\text{cosine}}(z) = \frac{1}{2}\sum_{k=1}^{K} g_k\left(\frac{1}{1-r_k e^{j\theta_k}z^{-1}} + \frac{1}{1-r_k e^{-j\theta_k}z^{-1}}\right) \quad (7)$$

Similarly, the impulse response of $H_{\text{cosine}}(z)$ consists of a sum of K damped cosines.

A different method is to use a simple all-pole structure to implement the filter:

$$H_{\text{all-pole}}(z) = \frac{1}{A(z)} = \frac{1}{\prod_{k=1}^{K} A_k(z)} \quad (8)$$

In this case, $B(z) = 1$ and the $B_k(z)$ can be retrieved by partial fraction expansion. This leads to a phase configuration strictly determined by the parameters of the denominator.

For this study, we would like to design a filter with a magnitude response that has a relatively modest dynamic range and also minimizes the depth of the valleys between peaks. We will make these notions more precise in Section 5.1, but let us qualitatively consider Figure 1, which shows the magnitude responses of the three common filter structures modelling five equally-damped modes.

The all-pole structure leads to a high dynamic range—more than 100 dB over the entire spectrum and around 40 dB between

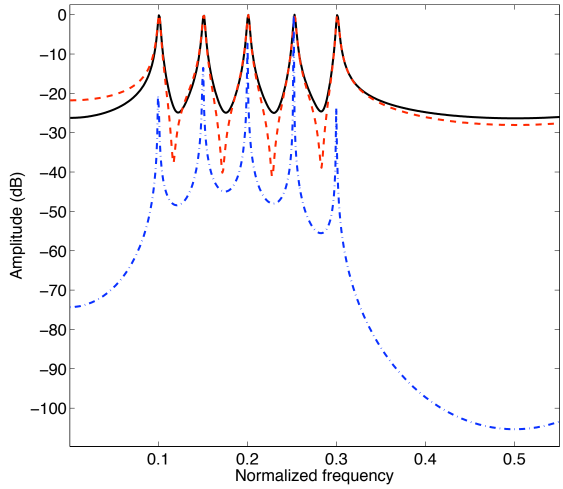

Figure 1: Magnitude responses of all-pole (dash-dotted line), sine section (dashed line) and cosine section (solid line) filter structures modelling five equally-damped modes.

the first and last peak. Note also that, although the modes are equally damped, the magnitudes of the peaks are not equal.

The parallel sine structure displays a much smaller dynamic range, and also spreads energy equally over the modes. Unfortunately, the valleys between the modes are very deep and sharp; this will likely result in artifacts during inverse filtering.

The parallel cosine structure compares favourably by achieving the lowest dynamic range, smooth valleys between the poles, and a good spread of energy over the modes. This structure has been considered in [3]. The authors concluded that when considering real cases with closely-spaced modes, the antiresonances were still too strong, leading to the presence of artifacts in the excitation signal.

Noticing that the antiresonance properties of sine and cosine section filters are determined by the global phase configuration, this lead us to the conclusion that the antiresonance phenomenon is created by the phase relationships between the second-order structures $\frac{B_k(z)}{A_k(z)}$. To mitigate the effects of antiresonances further, we propose to control the zeros of $B_k(z)$ by controlling the initial phase ϕ_k of each mode k independently using equation (2).

3. PROPOSED APPROACH

In order to determine the ϕ_k's such that the magnitude response valleys are as smooth as possible, we ideally require the roots of the numerator of (3). For anything but the simplest filters, numerical issues make this computation impossible in practice. Consequently, we propose to assume that only the poles immediately adjacent to a valley in the magnitude response have any appreciable effect on antiresonance properties.

Let us therefore consider one-pole filters $H_k(z)$ of the form

$$H_k(z) = \frac{g_k e^{j\phi_k}}{1-r_k e^{j\theta_k}z^{-1}}. \quad (9)$$

To control the magnitude of the valley between two such poles, we

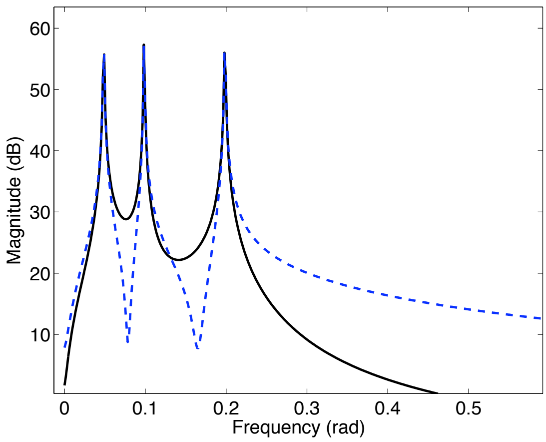

Figure 2: Magnitude responses of cosine section filter (dashed line) and IPA filter (solid line) modelling synthetic set of modes.

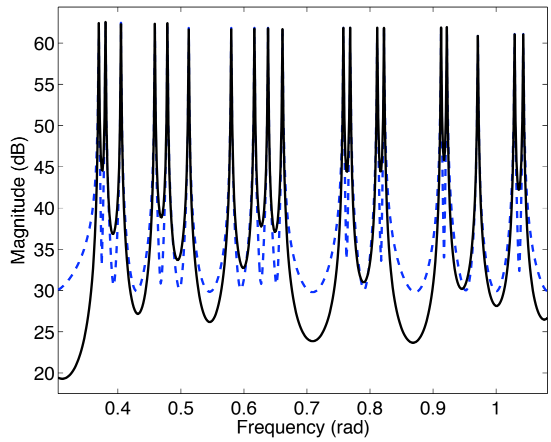

Figure 3: Magnitude responses of cosine section filter (dashed line) and IPA filter (solid line) modelling a set modes extracted from a percussive sound. The frequency axis is constrained to 0.3 to 1.1 radians.

propose to find the phase difference between the two poles such that the interference between the two poles over the bottom of the valley is minimized. Assume that we know the angular frequency $\omega_{\min}$ that minimizes the magnitude response in the valley between two poles $H_k(z)$ and $H_{k+1}(z)$ when $\phi_k = \phi_{k+1} = 0$, that is

$$\omega_{\min} = \underset{\omega\in(\theta_k,\theta_{k+1})}{\operatorname{argmin}} \left| \frac{g_k}{1 - r_k e^{j\theta_k} e^{-j\omega}} + \frac{g_{k+1}}{1 - r_{k+1} e^{j\theta_{k+1}} e^{-j\omega}} \right|. \quad (10)$$

This frequency can be approximated using the derivation of Section 4.

The phase response of $H_k(z)$ is given by

$$\angle H_k(e^{j\omega}) = \phi_k - \arctan\left[\frac{r_k \sin(\omega - \theta_k)}{1 - r_k \cos(\omega - \theta_k)}\right] \quad (11)$$

For a given value of ϕ_k, ϕ_{k+1} can be set such that the relationship between the poles is neither constructive nor destructive, i.e.:

$$\angle H_{k+1}(e^{j\omega_{\min}}) = \angle H_k(e^{j\omega_{\min}}) + \frac{\pi}{2} \quad (12)$$

The proposed method, which we call incremental phase adaptation (IPA), is used to incrementally set the phases of the poles. The algorithm is initialized by arbitrarily setting $\phi_1 = 0$. The initial phase ϕ_2 of the next pole is then computed using equation (12). The poles are assumed to be sorted in order of increasing frequency θ_k.

Figures 2 and 3 compare the magnitude responses of filters designed using IPA to cosine section filters created using equation (7).

4. FREQUENCY AT THE BOTTOM OF A VALLEY

A reasonable assumption to find an approximation of $\omega_{\min}$ is to solve

$$\left| \frac{g_k}{1 - r_k e^{j\theta_k} e^{-j\omega_{\min}}} \right| = \left| \frac{g_{k+1}}{1 - r_{k+1} e^{j\theta_{k+1}} e^{-j\omega_{\min}}} \right|. \quad (13)$$

After some algebra, we have

$$\omega_{\min} = \arctan(b_k, a_k) \pm \arccos\left[\frac{g_{k+1}^2(1+r_k^2) - g_k^2(1+r_{k+1}^2)}{\sqrt{a_k^2 + b_k^2}}\right],$$

where

$$a_k = 2[g_k^2 r_{k+1}\cos(\theta_{k+1}) - g_{k+1}^2 r_k \cos(\theta_k)]$$
$$b_k = 2[g_k^2 r_{k+1}\sin(\theta_{k+1}) - g_{k+1}^2 r_k \sin(\theta_k)].$$

For real solutions we must have

$$\left| g_{k+1}^2(1+r_k^2) - g_k^2(1+r_{k+1}^2) \right| \le \sqrt{a_k^2 + b_k^2}.$$

Negative frequency solutions must be offset by 2π to ensure $\omega_{\min} \in [0, 2\pi)$, and we simply choose the $\omega_{\min} \in (\theta_k, \theta_{k+1})$ as the appropriate value.

5. EXPERIMENTS

In this section, the proposed IPA filter design scheme is compared with the all-pole, sine section, and cosine section designs. Some criteria will be considered to evaluate the performance of the different approaches using databases of synthetic and real modes to test the designs.

5.1. Performance Criteria

We formulate three criteria to evaluate a filter $H(z)$ modelling a resonator with K modes at angular frequencies θ_k: the overall dynamic range (ODR), the pole dynamic range (PDR), and the average valley curvature (AVC).

We denote the magnitude response of the filter $H(z)$ by $G(\omega) = 20\log_{10}|H(e^{j\omega})|$. The ODR is then defined as the difference between the maximum and minimum values of the magnitude response:

$$\text{ODR}\{G\} = \max_{\omega\in[-\pi,\pi)}[G(\omega)] - \min_{\omega\in[-\pi,\pi)}[G(\omega)]. \quad (14)$$

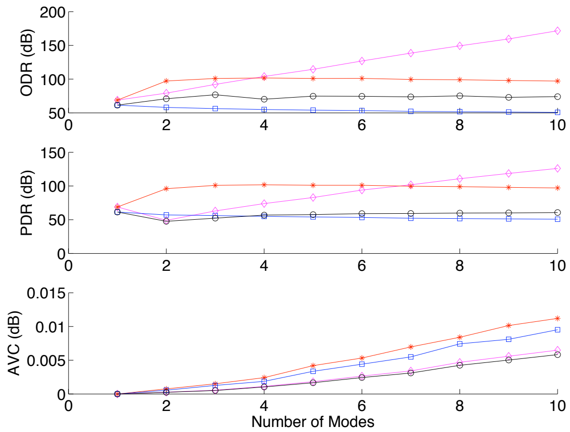

Figure 4: ODR, PDR, and AVC versus the number of modes considered for the four different methods of filter design: all-pole (diamonds), sine sections (stars), cosine sections (squares), and IPA (circles).

The PDR is defined as the difference between the maximum and minimum peaks of the filter magnitude response:

$$\mathrm{PDR}\{G\} = \max_{\omega \in \{\theta_1,\ldots,\theta_K\}} [G(\omega)] - \min_{\omega \in \{\theta_1,\ldots,\theta_K\}} [G(\omega)]. \quad (15)$$

The AVC is defined as

$$\mathrm{AVC}\{G\} = \frac{1}{K-1}\sum_{k=1}^{K-1} \frac{1}{\theta_{k+1}-\theta_k-2\varepsilon}\int_{\theta_k+\varepsilon}^{\theta_{k+1}-\varepsilon} G''(\omega)\,d\omega \quad (16)$$

$$= \frac{1}{K-1}\sum_{k=1}^{K-1} \frac{G'(\theta_{k+1}-\varepsilon)-G'(\theta_k+\varepsilon)}{\theta_{k+1}-\theta_k-2\varepsilon}. \quad (17)$$

Here, ε is a positive number used to discard the effects of peaks on the AVC.

5.2. Synthetic Data

In the first experiment, we compute filters modelling sets of $K \in [1,20]$ modes, with mode frequencies θ_k and damping parameters $d_k = \ln(r_k)$ randomly selected such that $\theta_k \in (0,\pi)$ radians and $d_k \in (0,0.001)$. We compute 1,000 filters for each set of K modes and take the mean values of ODR, PDR, and AVC over the filters.

Figure 4 compares the performance of filters designed using all-pole sections, sine sections, cosine sections, and IPA. (The trends in performance do not change beyond ten modes.) The ODR and PDR curves for the standard filter structures confirm the qualitative discussion of Section 2. Filters designed using IPA have a higher dynamic range (both in terms of ODR and PDR) than filters designed using cosine sections. This appears to be the price of the significant improvement in AVC.

5.3. Percussive Data

In this experiment, we considered 147 sets of modes extracted using high resolution approaches [7] from a database of percussive sounds recorded by Bruno Giordano[1]. Details about the type of hammers and plates, as well as the recording settings, can be found in [5]. The results displayed in Table 1 globally confirm the results obtained in the synthetic case.

[1]Results of deconvolution using the three different filter designs will be presented at the conference.

	ap	sin	cos	IPA
ODR (dB)	0.010	0.009	0.005	0.007
PDR (dB)	0.008	0.009	0.005	0.005
AVC	4.92e-07	7.52e-07	6.32e-07	4.17e-07

Table 1: ODR, PDR, and AVC criteria for the four methods of filter design over sets of modes extracted from percussive sounds: all-pole (ap), sine sections (sin), cosine sections (cos), and incremental phase adaptation (IPA).

6. ACKNOWLEDGMENTS

The authors would like to thank Bruno Giordano for letting us use his database of sounds, and Roland Badeau for kindly providing us with the high resolution analysis method used to estimate the modal parameters. This work was supported by a Special Research Opportunity grant from the Natural Sciences and Engineering Research Council of Canada.

7. CONCLUSION

The source-filter model is a very convenient approach for the modelling of a variety of natural phenomena. One of the major issues is the estimation of the excitation signal since only the filter output is available in common recording settings. The proposed approach attempts to improve the inverse filtering process by individually controlling the phase of the different modes to reduce some artifacts that limit the applicability of this process.

8. REFERENCES

[1] Steven M. Kay, *Modern Spectral Estimation: Theory and Application*, pp. 228–231, Prentice Hall, Upper Saddle River, New Jersey, 1988.

[2] Jordi Bonada, "High quality voice transformations based on modeling radiated voice pulses in frequency domain," in *Proceedings of the 7th International Conference on Digital Audio Effects*, Naples, Italy, Oct. 2004, pp. 291–295.

[3] Jean Laroche and Jean-Louis Meillier, "Multichannel excitation/filter modeling of percussive sounds with application to the piano," *IEEE Transactions on Speech and Audio Processing*, vol. 2, pp. 329–344, Apr. 1994.

[4] Jean-Louis Meillier, *Analyse/Synthèse de signaux percussifs par un modèle source/filtre*, Ph.D. thesis, Université du Maine, Le Mans, France, Feb. 1993, in French.

[5] Bruno L. Giordano, *Sound Source Perception in Impact Sounds*, Ph.D. thesis, University of Padua, Italy, June 2005.

[6] Xavier Serra, "Musical sound modeling with sinusoids plus noise," in *Musical Signal Processing*, G. De Poli, A. Piccialli, S. T. Pope, and C. Roads, Eds., pp. 91–122. Swets & Zeitlinger, Lisse, the Netherlands, 1997.

[7] Roland Badeau, Gaël Richard, and Bertrand David, "Fast adaptive ESPRIT algorithm," in *Proceedings of the IEEE/SP 13th Workshop on Statistical Signal Processing*, Bordeaux, France, July 2005, pp. 289–294.