Publication III

Heidi-Maria Lehtonen, Vesa Välimäki, and Timo I. Laakso. 2008. Canceling and selecting partials from musical tones using fractional-delay filters. Computer Music Journal, volume 32, number 2, pages 43-56.

Canceling and Selecting Partials from Musical Tones Using Fractional-Delay Filters

Heidi-Maria Lehtonen,* Vesa Välimäki,* and Timo I. Laakso*†
*Helsinki University of Technology (TKK)
Department of Signal Processing and Acoustics
P.O. Box 3000, FI-02015 TKK, Espoo, Finland
†National Board of Patents and Registration of Finland
Patents and Innovations Line
P.O. Box 1140, FI-00101, Helsinki, Finland
{heidi-maria.lehtonen, vesa.valimaki, timo.laakso}@tkk.fi

This article discusses canceling and extracting harmonics from a musical signal using digital filters. This is an old technique that has been proposed in different forms by Moorer in the 1970s for pitch detection of speech signals (Moorer 1974) and for analyzing music data for additive synthesis (Moorer 1977). The basic idea is to use a multiple-notch filter to extract individual harmonic components as signals. The filter structure can be obtained as the inverse transfer function of a comb filter (i.e., a delay line in a feedback loop).

In this article, we expand on a recently proposed idea that the delay line in the inverse comb filter (ICF) can be replaced with a high-order fractional-delay filter to obtain very accurate cancellation of neighboring harmonics to select a single harmonic or to extract the residual signal by canceling all harmonics (Välimäki, Ilmoniemi, and Huotilainen 2004; Välimäki, Lehtonen, and Laakso 2007). The proposed signal analysis method is useful for many practical cases. Many musical instruments, including all woodwind, brass, and bowed string instruments, produce a sound signal that is inherently harmonic, that is, the spectral components are integral multiples of a fundamental frequency. This follows from the sound-production mechanism of these self-excited systems, which involves mode locking in the time domain (Fletcher and Rossing 1991). It forces the sustained tones of such instruments to be periodic. There is often a noise component in these musical tones, however, making them pseudo-periodic in practice.

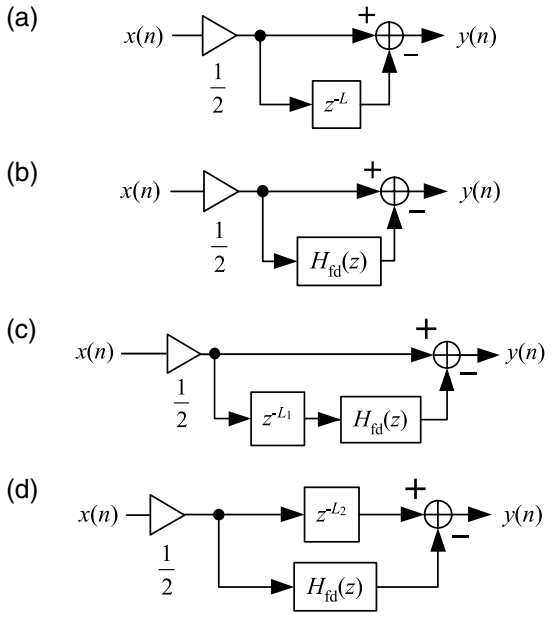
Another method for this kind of signal decomposition is sinusoidal modeling (McAulay and Quatieri 1986; Serra 1989; Serra and Smith 1990). In this method, the signal is analyzed using the windowed Fast Fourier transform (FFT), and the frequency and amplitude tracks are obtained by connecting data in the neighboring analysis frames. This approach has its roots in the phase vocoder technique and its efficient transform-domain implementation. For periodic or pseudo-periodic musical tones it is unnecessary to resort to an overly generic analysis method, because the frequencies of the harmonic components are known after the estimation of the fundamental frequency. Advantages of the proposed filter-based analysis method (compared with the more general FFT-based techniques) are simplicity, which follows mainly from the small number of parameters, and the possibility of designing filter coefficients in closed form. Additionally, the resulting decomposition is obtained directly as a set of time-domain signals, and no separate synthesis stage is required.

Other signal processing methods proposed for analyzing the harmonic structure of musical signals include wavelets (Evangelista 1993) and high-resolution tracking methods (Badeau, David, and Richard 2006). These methods provide excellent frequency accuracy at the expense of a complicated algorithm and a high computational cost. The method proposed in this article can also provide amplitude and frequency accuracy that is sufficient for musical signal analysis, but at the same time, the analysis method remains easy to apply.

In this article, we first discuss the theory behind canceling and selecting partials using fractional delay filters. Then, we present three test cases to demonstrate the power of this approach in musical signal analysis. Sound samples corresponding to these examples are provided on the forthcoming *Computer Music Journal* DVD (to be released with the Winter 2008 issue).

Computer Music Journal, 32:2, pp. 43–56, Summer 2008

Figure 1. (a) Conventional ICF and (b) a fractional-delay filter based ICF (after Välimäki, Ilmoniemi, and Huotilainen 2004). Two other ICF structures are presented in (c) and (d) that allow the use of a fractional-delay filter $H_{fd}(z)$ of arbitrary order for signals with both (c) low and (d) high fundamental frequencies.



Fractional-Delay Inverse Comb Filters

The inverse comb filter (ICF) is a finite impulse response (FIR) filter in which input signal is delayed by L samples and is then subtracted from the original input signal (see Figure 1a). Following the convention of Steiglitz (1996), the term "inverse comb filter" is used for the feedforward system with a delay line. The "comb filter," in contrast, has a delay line inside a feedback loop. The transfer function of the inverse comb filter is $H_{\mathrm{ICF}}(z) = (1 - z^{-L})/2$, where the scaling factor 1/2 sets the gain to unity in the passband (i.e., between the notches). The magnitude response of this filter features periodic notches at the multiples of f_s/L, where f_s is the sampling rate (Hz) and L is the delay line length in samples or in multiples of the sampling interval.

When the delay line length is restricted to be an integral multiple of the sample interval, the accuracy of the notch frequencies can be poor. An example is shown in Figure 2 in which the fundamental frequency is 4,186 Hz and the corresponding period length is 10.5351 samples. The sample rate used is 44.1 kHz. Practical ICF implementations employ a fractional-delay filter that replaces the delay line (Laakso et al. 1996; Pei and Tseng 1998; Välimäki and Laakso 2001). Alternatively, an FIR (Dutta Roy, Jain, and Kumar 1994) or an infinite impulse response (IIR) notch filter (Pei and Tseng 1997; Tseng and Pei 2001) can be designed to approximate the overall ICF characteristics.

Figure 2. Magnitude response of the conventional (dashed line) and the fractional-delay all-pass filter (solid line) ICF. The thick vertical lines indicate the harmonic frequencies to be canceled (f_0 = 4,186 Hz, f_s = 44,100 Hz).

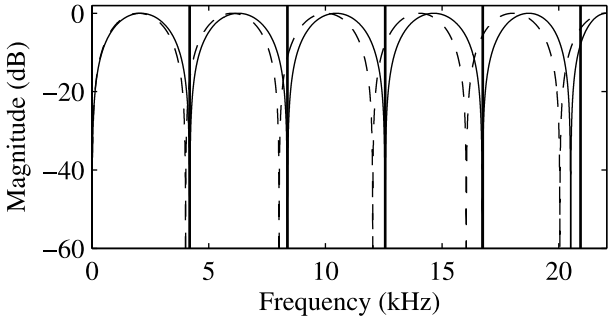


Figure 1b shows the block diagram of a fractional-delay ICF, where the delay is replaced with a fractional-delay filter, as proposed previously (Välimäki, Ilmoniemi, and Huotilainen 2004; Välimäki, Lehtonen, and Laakso 2007). The transfer function of this system can be written as $H(z) = (1 - H_{\mathrm{fd}}(z))/2$, where $H_{\mathrm{fd}}(z)$ is the transfer function of the fractional-delay filter used for delay approximation. A magnitude response of this structure with an 11th-order all-pass fractional-delay filter that approximates the delay of 10.5351 sampling intervals is displayed in Figure 2 (solid line).

The filter structures in Figures 1c and 1d offer freedom in the selection of the fractional-delay filter order (Välimäki, Ilmoniemi, and Huotilainen 2004). We have found experimentally that the order of $H_{\mathrm{fd}}(z)$ can be kept constant (e.g., N = 80), so when the fundamental period ($T_0 = f_s/f_0$) is longer than N samples, L_1 extra samples of delay are required in the lower signal path in Figure 1c. However, when the fundamental period is shorter than N samples, L_2 extra samples are required in the upper signal path to synchronize signals for subtraction (see Figure 1d). Thus, we propose to use the transfer function

$$H_{\mathrm{low}}(z) = \frac{1}{2}\left[1 - z^{-L_1} H_{\mathrm{fd}}(z)\right] \qquad (1)$$

when the fundamental period T_0 is larger than (or about the same as) the filter order N, and the transfer function

$$H_{\text{high}}(z) = \frac{1}{2}\left[z^{-L_2} - H_{\text{fd}}(z)\right] \quad (2)$$

when the fundamental period T_0 is smaller than the filter order N. The delay-line lengths L_1 and L_2 are set to $L_1 = T_0 - N - d$ (when $T_0 \geq N$) and $L_2 = N + d - T_0$ (when $T_0 < N$). Here, $-1 < d < 1$ is the fractional-delay parameter used in designing the filter.

FIR Fractional-Delay Filter Design

Let us first consider the inverse comb filter design using FIR fractional-delay filters. For simplicity, we assume that the length of the delay element $L_1 = L_2 = 0$, namely, no additional delay is present in either branch. This corresponds to the situation depicted in Figure 1b, and the transfer function can be written as

$$H(z) = \frac{1}{2}\left[1 - H_{\text{fd}}(z)\right] \quad (3)$$

where the Nth-order transfer function $H_{\text{fd}}(z)$ can be expressed as

$$H_{\text{fd}}(z) = \sum_{n=0}^{N} h(n) z^{-n} \quad (4)$$

Several methods for computing the FIR filter coefficients $h(n)$ exist (Laakso et al. 1996). Two commonly used methods are the windowed-sinc interpolation and the Lagrange interpolation. The impulse response, and thus the corresponding FIR filter coefficients, for the windowed-sinc interpolation can be expressed as

$$h_{\text{sinc}}(n) = \begin{cases} w(n-D)\,\text{sinc}(n-D) & \text{for } 0 \leq n \leq N \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where D is the desired delay in samples. The window function $w(n)$ (e.g., a Hamming window) is used for reducing the Gibbs phenomenon.

The Lagrange design formula, which can be expressed in closed form according to Laakso et al. (1996) as

$$h_{\text{L}}(n) = \prod_{\substack{k=0 \\ k \neq n}}^{N} \frac{D-k}{n-k} \quad \text{for } n = 0,1,2,\ldots,N \quad (6)$$

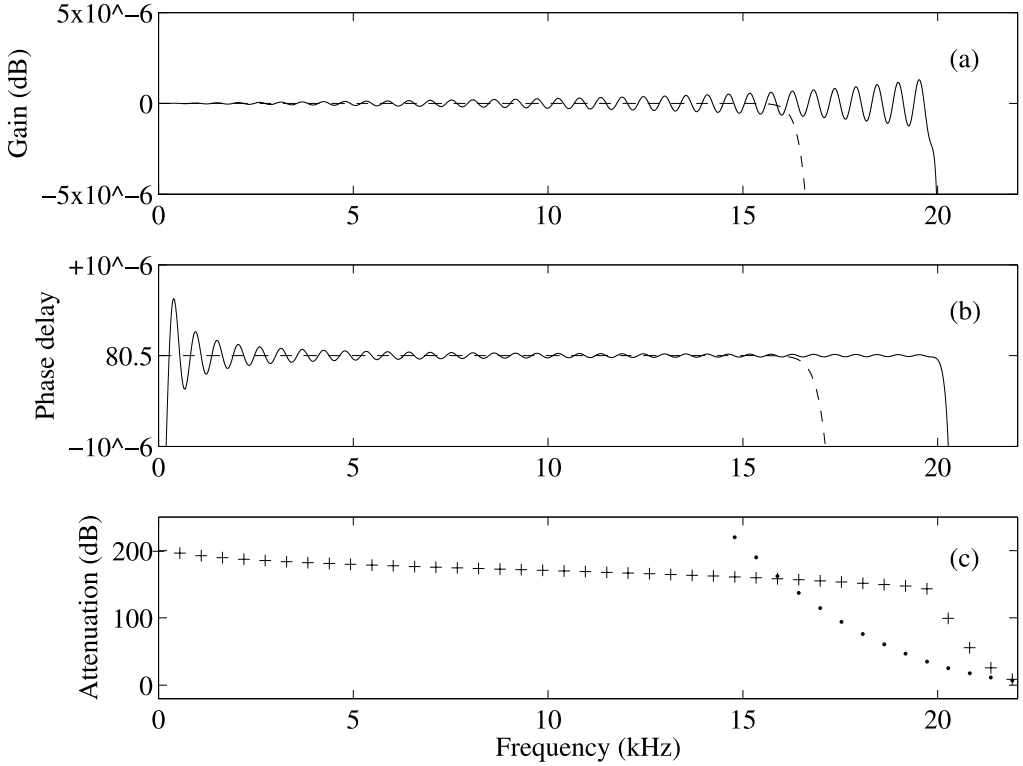
provides a maximally flat FIR fractional-delay filter design. Recently, Välimäki and Haghparast (2007) presented a design method in which the filter coefficients are obtained by truncating the impulse response (Equation 6) symmetrically from both ends. The explicit design formula for the truncated Lagrange design method can be written as

$$h_{\text{TL}}(n) = \prod_{\substack{k=0 \\ k \neq n+K_1}}^{M} \frac{D-k}{n+K_1-k} \quad \text{for } n = 0,1,2,\ldots,N \quad (7)$$

where M is the prototype filter order ($M > N$), and K_1 represents the number of coefficients that are discarded from either end of the impulse response (Equation 6). The advantage of the truncated Lagrange fractional-delay filter is the wider approximation bandwidth. This comes at the expense of increased ripple in the filter's frequency response.

Figure 3 compares the standard and the truncated Lagrange FIR filters. The filter order N is chosen to be 160 for both filters and the order of the prototype filter order M for the truncated Lagrange filter is set to $7N = 1120$. The fractional delay parameter d is 0.5. Figure 3a illustrates the magnitude responses of the two filters; note that the pass-band of the truncated Lagrange filter is wider than that of the standard filter. On the other hand, the ripple on the pass-band is larger. The same phenomenon occurs in the phase delay (i.e., the negative phase function divided by angular frequency) characteristics, which are visible in Figure 3b: The standard Lagrange filter provides a steadier response, but the approximation bandwidth is narrower compared to the truncated Lagrange filter. Figure 3c illustrates a comparison between two ICFs based on the standard and the truncated Lagrange FIR filters. For clarity, the lengths of the additional delay-lines were chosen to be $L_1 = L_2 = 0$. The dots and plus signs indicate the attenuation at the harmonic frequencies when an ICF with the standard and the truncated Lagrange FIR filters are used, respectively. Note that with the truncated Lagrange filter it is possible to obtain a 140 dB attenuation up to 20 kHz, whereas the per-

Figure 3. (a) Magnitude response, (b) phase delay in samples, and (c) attenuation of harmonics of ICFs using standard and truncated Lagrange FIR filters. The standard Lagrange FIR filter is indicated with a dashed line in (a) and (b) and with a dot at each attenuated harmonic in (c). The truncated Lagrange FIR filter is indicated with a solid line in (a) and (b) and with plus signs in (c). The parameter values are N = *160,* M = *1120, and* d = *0.5.*



formance of the standard Lagrange filter collapses above 16 kHz.

All-Pass Fractional-Delay Filter Design

For all-pass fractional-delay filter design, we can express the transfer function (Equation 1) in the form

$$H(z) = \frac{1}{2}\left[1 - H_{\text{fd}}(z)\right] = \frac{1}{2}\left[1 - \frac{z^{-N}D(z^{-1})}{D(z)}\right] = \frac{1}{2}\left[\frac{D(z) - z^{-N}D(z^{-1})}{D(z)}\right] \quad (8)$$

where $D(z) = 1 + a(1)z^{-1} + a(2)z^{-2} + \ldots + a(N)z^{-N}$ and the numerator polynomial can be written as $B(z) = D(z) - z^{-N}D(z^{-1}) = b(0) + b(1)z^{-1} + \ldots + b(N)z^{-N}$, and $b(k) = a(k) - a(N–k)$, $k = 0, 1, \ldots, N$.

It is easy to verify that $B(z)$ is an *antisymmetric* polynomial, that is, $b_k = –b_{N-k}$. In fact, the ICF is a special case of the parallel connection of two all-pass filters discussed by Saramäki (1985). As shown in that seminal paper, under general assumptions (stable all-pass functions) the overall numerator polynomial of this structure is either symmetric or antisymmetric (a mirror-image or anti-mirror-image polynomial, respectively). These polynomials are known to have their zeroes exactly on the unit circle. Hence, our filter is known to have accurate zeroes also in the fractional-delay ICF case. Note that the all-pass filters are exactly all-pass (unity magnitude at all frequencies) even if the phase (or phase delay or group delay) is only approximately as desired. In particular, as shown by Välimäki, Lehtonen, and Laakso (2007), the phase approximation errors of the all-pass filter tend to accumulate with increasing frequency, and the zeroes of the corresponding ICF further from the ideal places at higher frequencies.

For good accuracy, a method is needed that allows for the design of high-order fractional-delay filters. Three closed-form design methods are known for fractional-delay all-pass filters: the Thiran all-pass filter design (Thiran 1971; Fettweis 1972; Laakso et al. 1996), the truncated Thiran all-pass filter

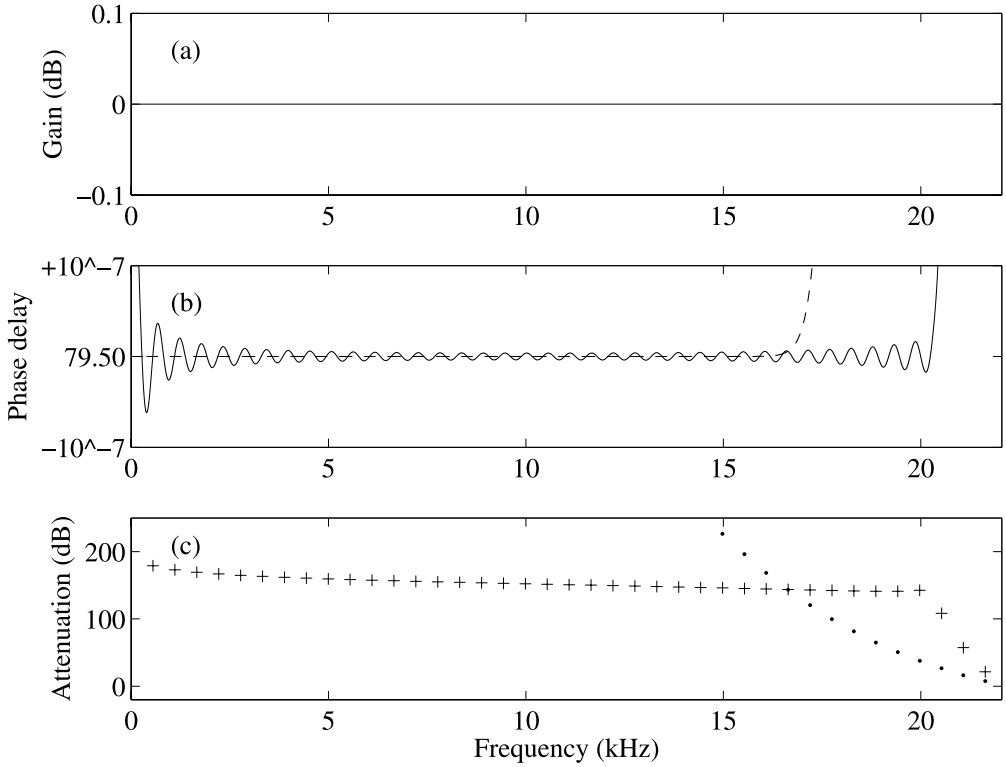


Figure 4. (a) Magnitude response, (b) phase delay in samples, and (c) attenuation of harmonics of ICFs using standard and truncated Thiran all-pass filters. The standard Thiran all-pass filter is indicated with a dashed line in (a) and (b) and with a dot at each attenuated harmonic in (c). The truncated Thiran all-pass filter is indicated with a solid line in (a) and (b) and with plus signs in (c). The parameter values are N = *80,* M = *720, and* d = *–0.5.*

(Välimäki 2000), and the Pei-Wang method (Pei and Wang 2004). Such methods are needed to increase the all-pass filter order to be large enough for good wideband approximation in audio applications. Both the standard and the truncated Thiran methods allow the filter order to be increased up to N = 1,029 (when d = –0.5) using 64-bit double floating-point computing. An alternative method would be the non-parametric all-pass filter design method proposed by Abel and Smith (2006). This method provides numerical stability even with large filter orders, and it is possible to match an arbitrary group delay as a function of frequency.

In this work, we concentrate on the standard and the truncated Thiran filter design methods. The Thiran design formula for the coefficients of the polynomial $D(z)$ in Equation 8 can be expressed as

$$a(k) = (-1)^k \binom{N}{k} \prod_{n=0}^{N} \frac{d+n}{d+k+n} \text{ for } k = 1,2,3,\ldots,N \quad (9)$$

where N is the filter order, and d is the fractional delay parameter ($-0.5 < d \leq 0.5$) (Laakso et al. 1996). At low frequencies, this filter has a phase delay of $N + d$ samples.

The truncated Thiran design is obtained by modifying Equation 9:

$$a(k) = (-1)^k \binom{M}{k} \prod_{n=0}^{M} \frac{d+n}{d+k+n} \text{ for } k = 1,2,3,\ldots,N \quad (10)$$

where M is the prototype filter order ($M > N$) (Välimäki 2000). By using a value for M that is much larger than N in Equation 10, it is possible to extend the bandwidth of good approximation. This comes at the expense of losing quality at low frequencies: The approximation error is larger than in the original all-pass filter. This design technique allows a useful tradeoff between approximation accuracy and bandwidth, as discussed in Välimäki (2000); Välimäki, Ilmoniemi, and Huotilainen (2004); and Välimäki, Lehtonen, and Laakso (2007).

Figure 4 compares the standard and truncated Thiran all-pass filters. The design parameters are N = 80 and d = –0.5 for both filters, and the proto-

type filter order for the truncated Thiran filter is $M = 9N = 720$. The magnitude response, which is exactly flat in both cases, and the phase delays are displayed in Figures 4a and 4b, respectively.

Note from Figure 4b that the difference between the fractional-delay approximations of the two filters is microscopic below about 17 kHz, but the relaxed accuracy allows for the truncated Thiran filter to perform significantly better above 17 kHz. A comparison of the two ICFs with Thiran and truncated Thiran filters is shown in Figure 4c. Again, the lengths of the additional delay-lines were chosen to be $L_1 = L_2 = 0$. Note that below 17 kHz, the ICF with the truncated Thiran filter is worse than the ICF using the standard Thiran filter, but both are sufficiently good, because the attenuation is more than 140 dB. Above 17 kHz, the performance of the Thiran ICF collapses, but with the truncated version of the all-pass filter the ICF offers an attenuation of 140 dB up to 20 kHz.

Extracting Harmonic Components

Instead of canceling all the harmonic components, single harmonics can be selected and the rest canceled. This is achieved by cascading with an ICF a second-order all-pole filter that cancels a zero at a given harmonic frequency. This section describes the design of such a filter, which we call the *harmonic-extraction filter* (HEF).

It is not recommended to place a pole exactly on the unit circle in the z plane, because the resulting second-order filter is marginally stable and the hidden pole may cause numerical problems. A better approach is to move the zeroes of the ICF slightly inside the unit circle by defining the radius of all zeroes of the transfer function to be $1 - \varepsilon$, where ε is a very small non-negative constant. Consequently, the pole of the second-order filter can also have the same radius, so that the stability of the recursive filter can be assured.

To place all the zeroes at radius r, the coefficient of an ICF with a delay-line of length L must have a filter coefficient r^L (Orfanidis 1996). Consequently, a scaling coefficient

$$g_0 = 1/(1 + r^L) \tag{11}$$

must be used to ensure the maximum gain of the filter to be unity (i.e., 0 dB). Then, the minimum gain of the filter, which occurs at the bottom of the notches at harmonic frequencies, is $g_0(1 - r^L)$, which we call A. We can now solve for the required g_0, and consequently the required r, when gain A is set to a given value. By combining Equation 11 and the aforementioned definition for A, we get a representation for the radius r, as shown in Välimäki, Lehtonen, and Laakso (2007):

$$r = \sqrt[L]{(1 - A)/(1 + A)} \tag{12}$$

The HEF transfer function can be written as

$$H_{\text{HEF}}(z) = g_1 R(z)[1 - r^L H_{\text{fd}}(z)] \tag{13}$$

Here, the scaling coefficient g_1 that sets the maximum gain at the bottom of the notches (without the resonator) to be unity is defined by

$$g_1 = 1/\left[r(1 - r^L)\right] \tag{14}$$

and the transfer function of the resonant filter is

$$R(z) = b_0/(1 + a_1 z^{-1} + a_2 z^{-2}) \tag{15}$$

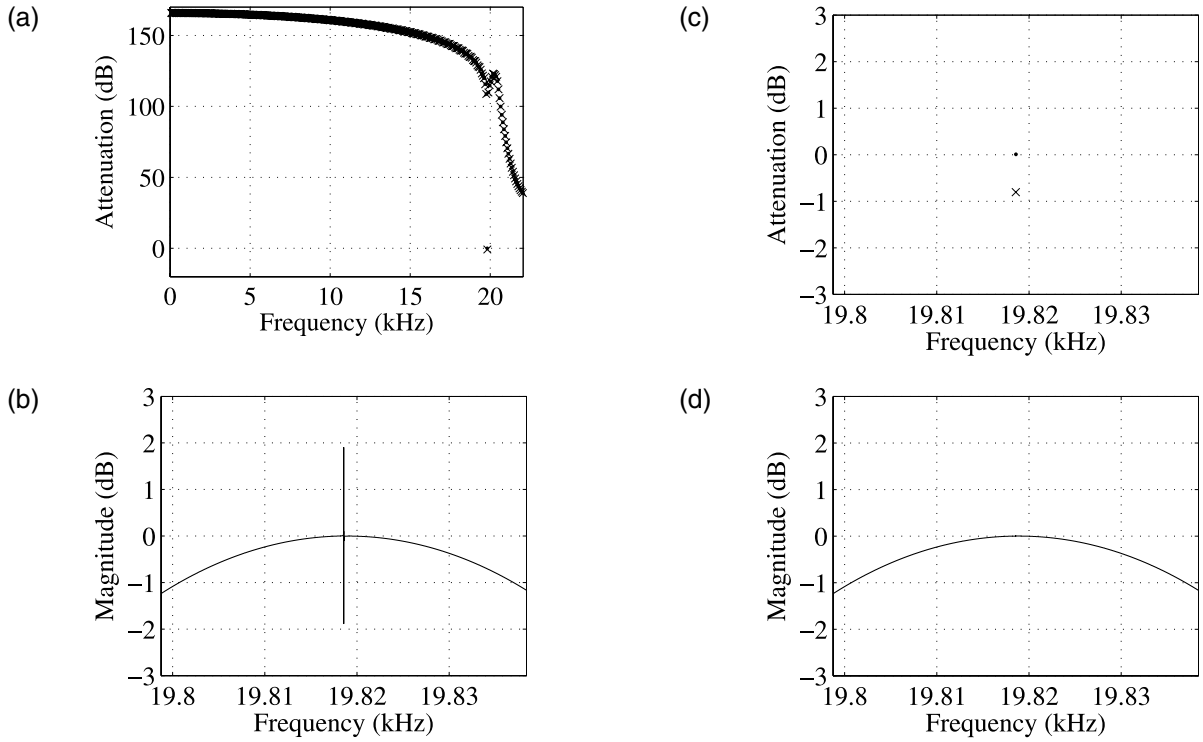
with coefficients $b_0 = (1 - r^2)\sin(2\pi f_{\text{res}}/f_s)$, which scales the maximum gain of the resonant filter to be unity (Steiglitz 1996), $a_1 = -2r\cos(2\pi f_{\text{res}}/f_s)$, and $a_2 = r^2$, where f_{res} is the resonant frequency that determines which harmonic component is retained. The filter that has the transfer function of Equation 13 with the given scaling coefficients has a maximum gain of 0 dB at the peak of the passband.

It should be pointed out that, because all the zeroes are inside the unit circle, the overall magnitude never reaches zero exactly, as shown by Välimäki, Lehtonen, and Laakso (2007). In practice, this is not a problem, because the achieved attenuation at the bottom of the notches is still very close to zero.

Design of Parameter Values

For good attenuation, it is required that A is sufficiently small and that the resonant filter $R(z)$ accurately cancels one of the zeroes of the ICF. For example, when it is required that the inverse comb filter attenuate the harmonic frequencies by 100 dB, the value of A must be set to 10^{-5}, because $20 \log_{10}(10^{-5}) = -100$ dB.

Figure 5. (a) Attenuation of harmonic partials using the single-harmonic canceling filter when the resonance frequency is the nominal mf_0 *('x'), and the corrected one ('.'). (c) The difference in magnitude with ('.') and without ('x') the correction around the partial of interest. Magnitude response of the HEF (b) without and (c) with the correction of the resonance frequency of the all-pole part.*



In practice, the high-order all-pass filter does not provide a perfect phase approximation. Thus, it may be necessary to set A to a smaller value, such as 10^{-6}. However, when the filter structure for selecting a single harmonic is used, the resonant filter provides additional attenuation at frequencies away from the resonance, which further improves the attenuation at the notches.

It was reported by Välimäki, Ilmoniemi, and Huotilainen (2004) that for some musical instrument tones with strong low-indexed harmonics, the filtering of the signal with the transfer function $H_{\mathrm{HEF}}(z)$ is insufficient. Listening to the filtered signal reveals that the fundamental is still perceived, although one of the high-frequency partials is strongly emphasized. Filtering the signal twice with a transfer function according to Equation 13 adequately attenuates the rest of the harmonics in this case, because it is then possible to obtain an attenuation that is better than 100 dB.

There is a minor mismatch in the cancellation of the mth transfer function zero with the pole of the resonant filter having the resonance frequency $f_{\mathrm{res}} = mf_0$, because the frequency of the mth zero is offset by the inaccuracy of the phase approximation of the all-pass filter. In practice, this mismatch produces a kink around the main lobe of the band-pass filter, and the gain at the resonance frequency becomes larger than 0 dB. A correction to the pole frequency is required to reduce this error.

One way to correct the resonance frequency of the all-pole filter is to search for the minimum of the ICF's magnitude response around the mth notch. For example, computing the magnitude response at 10,000 points between $0.999990mf_0$ and $1.000010mf_0$, and selecting f_{res} as the frequency, where the minimum occurs, reduces the mismatch sufficiently. To reduce the number of magnitude-response evaluations, the local minimum can be estimated by using interpolation, as suggested by Välimäki, Lehtonen, and Laakso (2007).

Figure 5a gives an example of the attenuation obtained without and with the proposed correction of the resonance frequency when the fundamental frequency is 69.2957 Hz, the harmonic #285 at 19749.2 Hz is selected, the all-pass filter orders used

are $N = 80$ and $M = 720$, and attenuation is $A = 10^{-5}$. In this case, the pole radius is $r = 1 - 31.4 \times 10^{-9} = 0.999999969$. The difference between the nominal (mf_0) and the corrected resonance frequency is 0.0135 Hz, but the attenuation of the partial is –0.8 dB without and 0.0045 dB with the correction. This difference is illustrated in Figure 5c. The difference in attenuation is enough to make the correction worth the effort, because it makes the analysis filter an accurate tool for signal analysis. Figures 5b and 5d display the magnitude response of the two filters around selected harmonic. It is seen that without correction, the gain of the filter fluctuates between about –2 dB and +2 dB near the resonance frequency. Using the correction, the fluctuation is negligible, not exceeding 0.010 dB (not visible in Figure 5d owing to limited image resolution; the kink is very small and very narrow).

For comparison, we designed two ICFs with the truncated Lagrange and truncated Thiran all-pass filters, and a linear-phase FIR band-pass filter that imitates the obtained magnitude response. The linear-phase FIR filter was designed using the Chebyshev window with a sidelobe level of –90 dB. The width of the pass-band was equal to f_0, that is, 69.2957 Hz. It was required that the magnitude error at the frequency of the selected harmonic not exceed 0.1 dB, and that the minimum attenuation at the frequencies of the canceled harmonics is 100 dB. Table 1 shows the results of the comparison in terms of the computational complexity when all filters are realized in direct form.

As can be seen in Table 1, the number of multiplications needed when the truncated Lagrange or truncated Thiran filter is used is only 3.1% of that of the FIR filter. The computational complexities of the truncated Lagrange and truncated Thiran filters are the same, but the orders of the prototype filters differ.

Table 1. Results of the Comparison Between the Linear-Phase FIR Filter, the Truncated Lagrange Filter, and the Truncated Thiran Filter in Terms of the Computational Complexity

Filter	*Order*	*Prototype order*	*Multiplications*	*Additions*
Linear-phase FIR	4246	—	4247	4246
Truncated Lagrange	130	911	131	130
Truncated Thiran	65	585	131	130

Separation of Odd and Even Partials

Although it is possible to cancel the harmonic components one by one by applying the above HEF structure multiple times, alternatively, even and odd harmonics can be separated using a single filtering operation, as suggested by Välimäki, Ilmoniemi, and Huotilainen (2004) and Välimäki, Lehtonen, and Laakso (2007). The odd and even harmonics can be separated by first canceling the even harmonics using the fractional-delay inverse comb filter and then subtracting the resulting signal from the original.

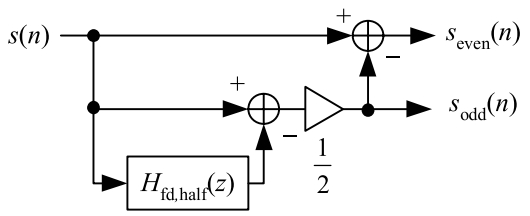
The structure of Figure 1c or 1d can be used, but the delay to be approximated is half of that used in canceling all harmonics, namely, $f_s/2f_0$ samples. With this delay, the notches are located at the multiples of the second harmonic, and the filter now cancels the even harmonics and preserves the odd harmonics. The signal containing even harmonics is then obtained by subtracting the estimated odd harmonics from the original signal, as shown in Figure 6:

$$s_{\text{even}}(n) = s_{\text{orig}}(n) - s_{\text{odd}}(n) \tag{16}$$

Case Studies

This section presents how the proposed filtering algorithms perform with synthetic signals and recorded instrument tones. In addition, the ICF structure based on either windowed-sinc interpolation, truncated Lagrange, or truncated Thiran interpolation are compared against sinusoidal modeling (McAulay and Quatieri 1986; Serra 1989;

Figure 6. Structure for separating the even and odd harmonics of a musical signal using one fractional-delay filter.



Serra and Smith 1990). The sampling frequency is 44.1 kHz in all the test cases.

Selecting a Harmonic from a Synthetic Test Signal

The synthetic test signal is determined to be the sum of sinusoids:

$$x(n) = A_{sc}(n)\sum_{k=1}^{K} \sin\left(\frac{2\pi nkf_0}{f_s} + \varphi_k\right) \qquad (17)$$

where $A_{sc}(n)$ is an envelope function, K is the number of harmonics present in the signal, f_0 is the fundamental frequency of the signal, and φ_k is the phase of the kth harmonic. In this case, the parameters were chosen as follows: $K = 84$, $f_0 = 261.626$ Hz (C4), which corresponds to the cycle length of 168.562 samples. The initial phases φ_k are uniformly distributed random numbers in the range $[0, 2\pi]$. To examine the temporal smearing resulting from the harmonic component extraction, the envelope of the signal is chosen to be rectangular, containing sharp transitions.

As an example, two components, the fundamental frequency component and the 76th overtone, have been extracted from the signal (Equation 17). This is done with the HEF structure (Equation 13) using one of the three fractional-delay filters: an FIR filter obtained with sinc interpolation using the Hamming window, a truncated Lagrange FIR filter, and a truncated Thiran IIR filter. The order of the FIR filters N was set to 160, and the prototype filter order for the truncated Lagrange filter was set to $M = 7N$. The order of the truncated Thiran filter was chosen to be $N = 80$, and the order of the prototype filter was set to $M = 9N$. The attenuation coefficient A was determined to be 10^{-6} in all cases. For comparison, the two components were also extracted using sinusoidal modeling, where the short-time FFT was computed using a Blackman window of length $4f_s/f_0$, and the FFT size was 2048. The hop size was set to be one-fourth of the window length.

The results are presented in Figures 7 and 8. Figures 7a and 7b present the spectrum of the original signal (Equation 17), and Figure 7c presents the spectrum of the partials #1 and #76 that are selected with an ICF based on a windowed-sinc interpolation. Figures 7d, 7e, and 7f present the corresponding result obtained with sinusoidal modeling, the truncated Lagrange FIR filter, and the truncated Thiran method, respectively. All spectra were calculated from a 0.5-sec excerpt taken between 0.1–0.6 sec of the signals. The Hamming window was used, and the spectra were computed using the discrete-time Fourier transform at 851 equally spaced points so that every 10th point matched one harmonic. This choice of parameters yields a clear visual representation of the sharp spectral peaks.

As can be seen in Figures 7e and 7f, the truncated Lagrange and truncated Thiran filters are able to extract the first harmonic (thick line) efficiently, and the other harmonics are properly attenuated. In the case of the 76th harmonic (thin line) component, the truncated Thiran filter seems to perform slightly better, but the difference is insignificant. The FIR filter obtained with windowed-sinc interpolation in Figure 7c does well with the first harmonic, but the error near the Nyquist limit is greater, and moreover, the low-pass nature of the magnitude response must be taken into account in the analysis to maintain the level of the original signal in high frequencies. The sinusoidal-modeling technique depicted in Figure 7d suffers from spectral leakage caused by the slightly inaccurate harmonic-frequency trajectory estimate. This problem could be facilitated with a longer window length in the signal-analysis phase, but this would cause more temporal smearing in the time domain. With this choice of window length, attenuation of neighboring harmonics is not as effective as with the other methods.

The effects of temporal smearing are illustrated in Figure 8. The signals are zoomed to the window 0.98–1.06 sec. Figures 8a and 8b show the original

Figure 7. Results of comparison in the frequency domain. Here, (a) and (b) represent the synthetic test signal. Selected harmonic components #1 (thick line near 0 kHz in (c)–(f)) and #76 (near 20 kHz) obtained with an ICF based on (c) windowed-sinc interpolation, (d) sinusoidal modeling, (e) truncated Lagrange interpolation, and (f) truncated Thiran interpolation.

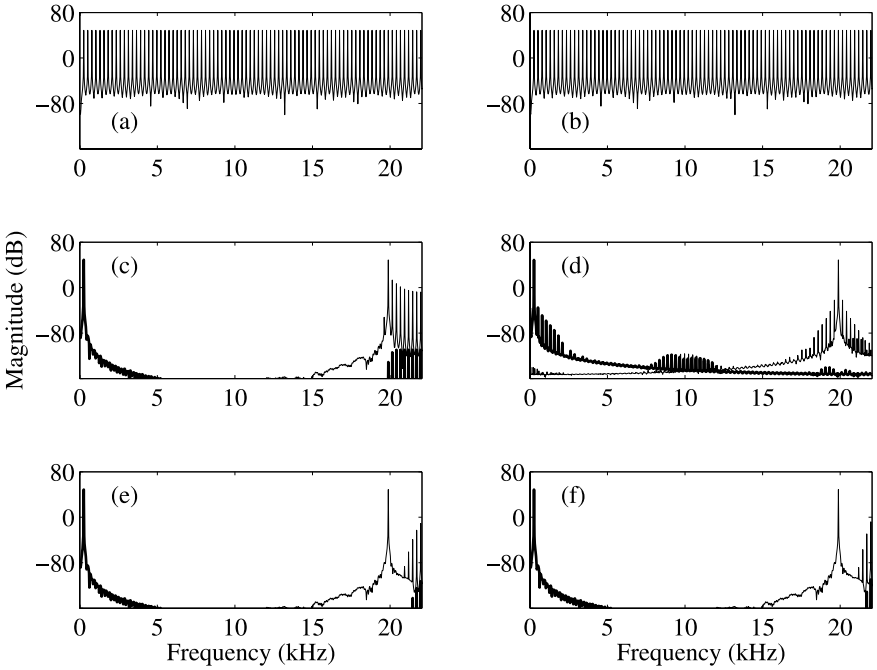


Figure 8. Effects of temporal smearing. Here, (a) and (b) show an excerpt of the original synthetic signal, and (c) and (d) show excerpts of the harmonics #1 and #76 obtained with an ICF based on the windowed-sinc interpolation. Also shown are corresponding signals obtained with (e)–(f) sinusoidal modeling; (g)–(h) a truncated Lagrange filter; and (i)–(j) a truncated Thiran filter.

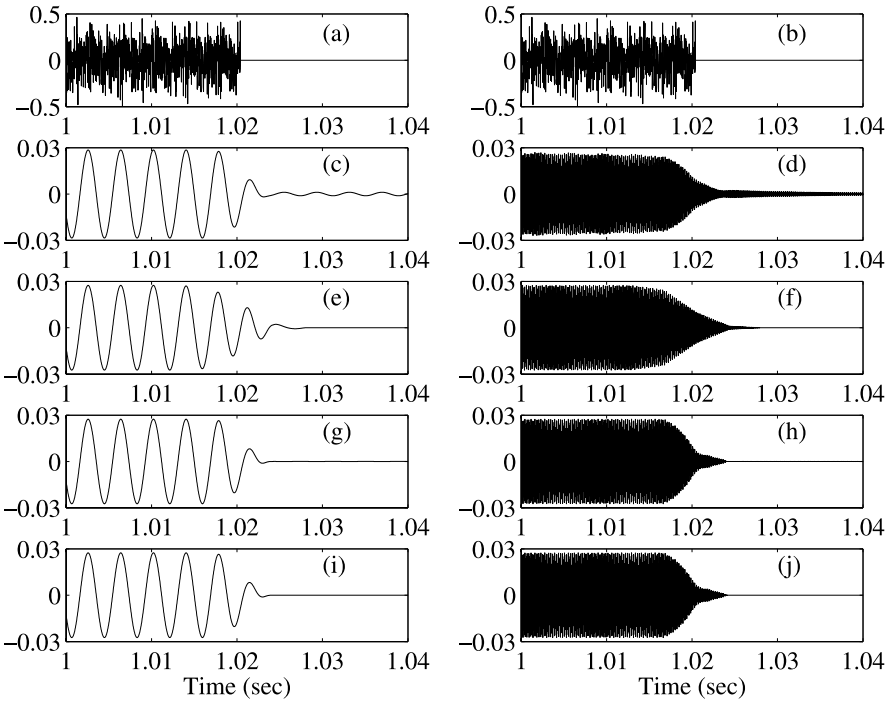
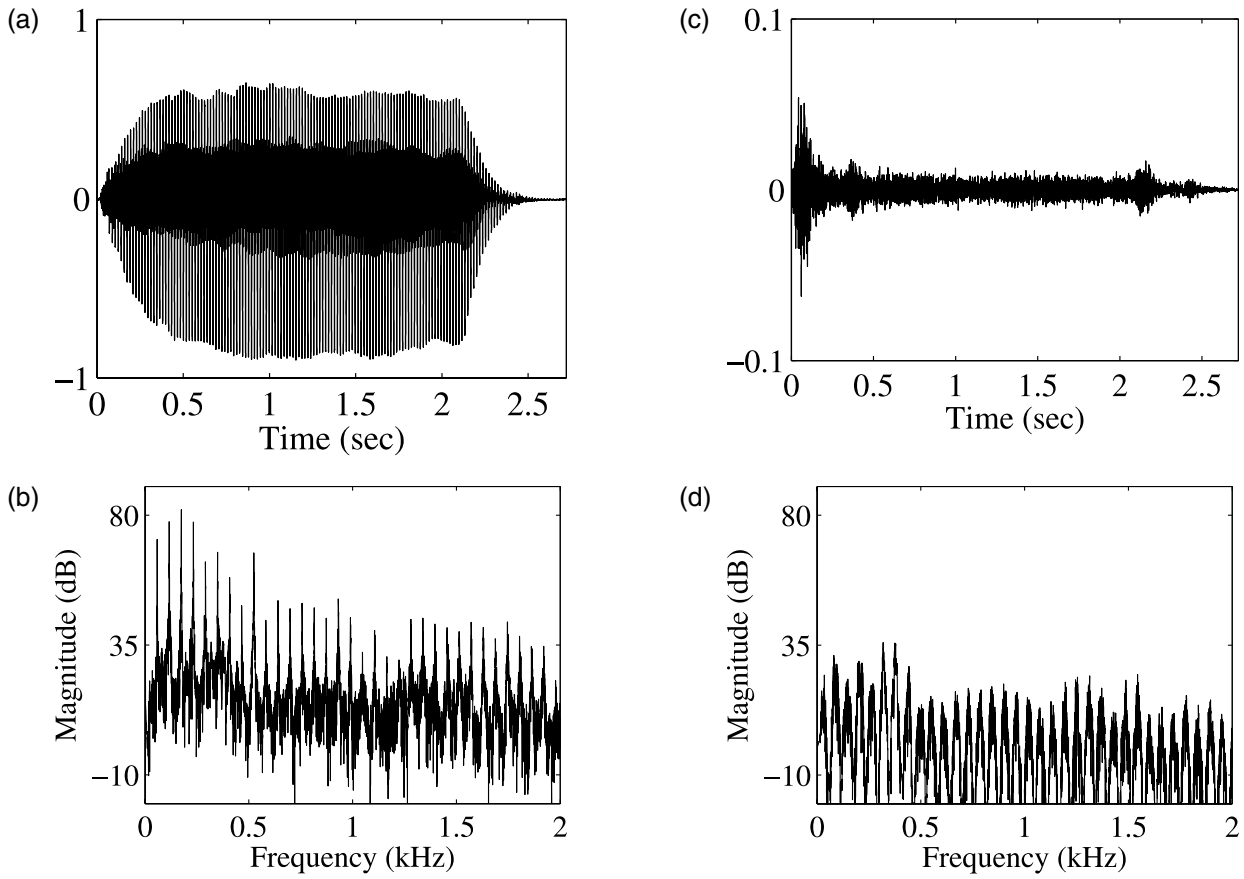


Figure 9. Time- and frequency-domain presentations of (a)–(b) the bowed double bass tone; (c)–(d) the extracted residual signal.



signal, and Figures 8c and 8d present the cases where the first harmonic and the 76th harmonic have been extracted with the ICF based on windowed-sinc interpolation. Temporal smearing in the sinusoidal modeling technique is depicted in Figures 8e and 8f. The performances of the truncated Lagrange FIR filter (see Figures 8g and 8h) and the truncated Thiran filter (see Figures 8i and 8j) are also shown. It can be seen that the truncated Lagrange and truncated Thiran filters perform best in terms of temporal smearing.

Residual Signal Extraction

To investigate how the ICF method works with recorded tones, a bowed double-bass tone was analyzed with an ICF based on the truncated Thiran filter. In Figures 9a and 9b, the tone is presented in the time and frequency domains, respectively. The fundamental frequency of the tone was measured to be $f_0 = 58.2306$ Hz using the YIN algorithm (de Cheveigné and Kawahara 2002).

It is now desired to remove all the harmonics instead of preserving one. Depending on the fundamental frequency, a modified form of the transfer function (Equation 3) can be used:

$$H(z) = g_1\left[1 - r^L H_{\text{fd}}(z)\right] \tag{18}$$

where the coefficients g_1 and r^L are computed with Equations 14 and 12, respectively. Again, the parameters were selected as follows: $N = 80$, $M = 9N$, and $A = 10^{-6}$. The filtered residual signal is presented in Figures 9c and 9d in time and frequency domains, respectively. When comparing Figures 9b and 9d, it can be seen that the harmonic

Figure 10. The structure of the kantele. The varras is magnified in order to illustrate the termination point. (Adapted from Pakarinen, Välimäki, and Karjalainen 2005.)

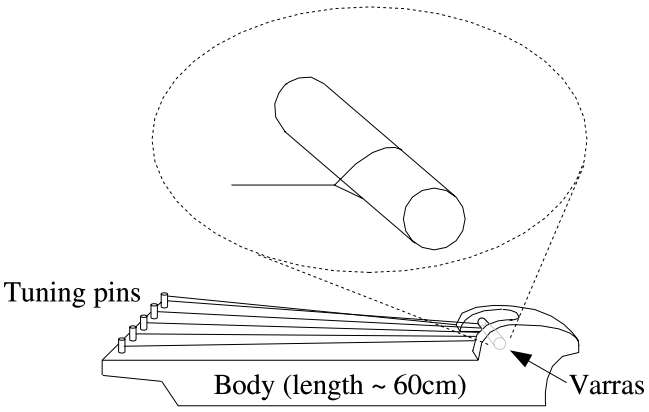
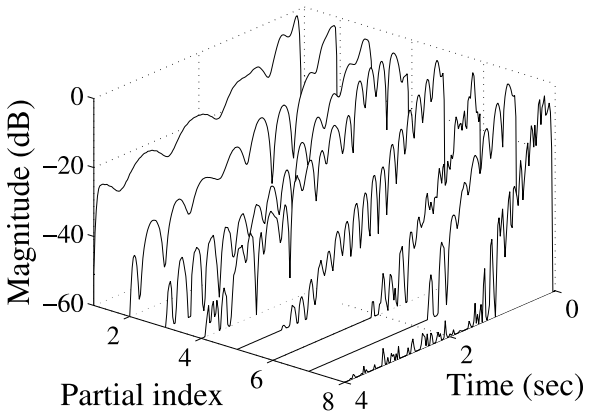


Figure 11. Envelopes of the first eight harmonics of a kantele tone.



components are attenuated efficiently. Moreover, the noise between the harmonics is preserved.

Extraction of a Beating Harmonic

Beating is a significant phenomenon present in many stringed instruments. It is characteristic to the piano and the *kantele,* for example. The kantele is a plucked string instrument that usually has five metal strings. The strings are terminated by metal tuning pins at one end and wound once around a horizontal metal bar called the *varras* at the other end and then knotted (see Figure 10). The beating in the kantele sound is caused by the fact that the vibrations in two polarization planes have different effective lengths, as the varras is the termination point for vibration in the horizontal plane and the knot is the termination point for the vibration in the vertical plane. This causes the tone to have two slightly different fundamental frequencies, which produces beating (Erkut et al. 2002).

Extraction of a beating harmonic is considered to be somewhat problematic with sinusoidal modeling, for example, because preservation of the beating in the resynthesized deterministic signal requires two or more spectral peaks to be extracted around a harmonic frequency (e.g., Tolonen 1999, and Esquef, Karjalainen, and Välimäki 2003). The beating harmonics of the kantele can be easily extracted from a recorded tone using the HEF structure (Equation 13).

Figure 11 shows a 3-D time-frequency plot of the envelopes of the first eight harmonics of a kantele tone. The dominating fundamental frequency of the tone was measured to be f_0 = 413.445 Hz. The truncated Thiran filter with parameters $N = 80$, $M = 9N$, and $A = 10^{-6}$ was used. In this case, the resulting attenuation is not quite 120 dB, owing to the two fundamental frequencies present in the signal. As the notches of the inverse comb filter are very sharp, part of the harmonic information does not hit the notch frequency but rather the slope of the notch, where the attenuation is less than 120 dB. The obtained attenuation is still more than 75 dB, which is sufficient for visualization, since the beating patterns are clear (see Figure 11). From these envelopes, it is easy to determine the frequency of the beating—for example, by fitting sinusoids to the envelopes and choosing the best match in the least-square sense (Erkut et al. 2002).

Conclusions and Future Work

Digital filtering techniques were proposed to obtain useful decompositions of harmonic musical signals. The basic approach taken here is to subtract a delayed copy of the signal from itself to cancel the harmonic components. A high-order fractional-delay filter implements an accurate approximation of the required time delay. A truncated Lagrange and a truncated Thiran all-pass fractional-delay filter are well suited for this task. A harmonic-

extraction filter is obtained by cascading a second-order all-pole filter with the inverse comb filter. Division of a musical signal into two signals—one containing the even and the other the odd harmonics—and the extraction of the background noise or residual were also suggested as promising operations that are easy to realize using the proposed filter structures.

Case studies were presented to show how the techniques perform in feature analysis for musical tones. Single harmonic components were extracted from a synthetic test tone. The neighboring harmonics were attenuated more than 100 dB. The residual noise component was extracted from a bowed string tone by suppressing all the harmonics. Finally, beating harmonics were extracted from a recorded kantele tone.

Future research includes developing a useful method to account for varying fundamental frequency of the signal, such as vibrato. In practice, this problem calls for a time-varying delay line to be used in the inverse comb filter. One suitable solution is to use a truncated Lagrange FIR filter (Välimäki and Haghparast 2007), which allows for accurate phase-delay approximation and real-time coefficient updating. A further application of the time-varying inverse comb filter would be the separation of simultaneous harmonic audio signals, such as musical polyphony or multiple people speaking at once, as discussed by de Cheveigné (1993).

Another interesting special case is the analysis of inharmonic tones, such as piano tones or other instrument sounds with regular inharmonicity caused by dispersion. A filter-based analysis tool for such tones requires an all-pass filter that approximates the dispersion in cascade with the delay line. This is a known method in digital waveguide synthesis of string sounds (e.g., Rocchesso and Scalcon 1996; Rauhala and Välimäki 2006). A more accurate approximation of dispersion characteristics is needed for the analysis tool than for sound synthesis. A possible choice would be to use the all-pass filter design method proposed by Abel and Smith (2006), because it allows the user to design the filter based on frequency-dependent group delay characteristics.

Acknowledgments

The work of Heidi-Maria Lehtonen has been supported by the GETA Graduate School, the Acadamy of Finland (project no. 122815), the Finnish Cultural Foundation, and the Nokia Foundation. The authors are grateful to Minna Ilmoniemi and Minna Huotilainen for their collaboration in the early stages of this work. Special thanks go to Jyri Pakarinen for giving us permission to use his fine drawing of the kantele.

References

Abel, J. S., and J. O. Smith. 2006. "Robust Design of Very High-Order All-Pass Dispersion Filters." *Proceedings of the 2006 International Conference on Digital Audio Effects.* Montreal: McGill University, pp. 13–18.

Badeau, R., B. David, and G. Richard. 2006. "High Resolution Spectral Analysis of Mixtures of Complex Exponentials Modulated by Polynomials." *IEEE Transactions on Signal Processing* 54(4):1341–1350.

de Cheveigné, A. 1993. "Separation of Concurrent Harmonic Sounds: Fundamental Frequency Estimation and a Time-Domain Cancellation Model of Auditory Processing." *Journal of the Acoustical Society of America* 93(6):3271–3290.

de Cheveigné, A., and H. Kawahara. 2002. "YIN, A Fundamental Frequency Estimator for Speech and Music." *Journal of the Acoustical Society of America* 111(4):1917–1930.

Dutta Roy, S. C., S. B. Jain, and B. Kumar. 1994. "Design of Digital FIR Notch Filters." *IEEE Proceedings of Vision, Image, and Signal Processing* 141(5):334–338.

Erkut, C., et al. 2002. "Acoustical Analysis and Model-Based Sound Synthesis of the Kantele." *Journal of the Acoustical Society of America* 112(4):1681–1691.

Esquef, P. A. A., M. Karjalainen, and V. Välimäki. 2003. "Frequency-Zooming ARMA Modeling for Analysis of Noisy String Instrument Tones." *EURASIP Journal on Applied Signal Processing* 2003(10):953–967.

Evangelista, G. 1993. "Pitch Synchronous Wavelet Representations of Speech and Music Signals." *IEEE Transactions on Signal Processing* 41(12):3313–3330.

Fettweis, A. 1972. "A Simple Design of Maximally Flat Delay Digital Filters." *IEEE Transactions on Audio and Electroacoustics* 20(2):112–114.

Fletcher, N. H., and T. D. Rossing. 1991. *The Physics of Musical Instruments.* New York: Springer.

Laakso, T. I., et al. 1996. "Splitting the Unit Delay—Tools for Fractional Delay Filter Design." *IEEE Signal Processing Magazine* 13(1):30–60.
McAulay, R. J., and T. F. Quatieri. 1986. "Speech Analysis/Synthesis Based on a Sinusoidal Representation." *IEEE Transactions on Acoustics, Speech, and Signal Processing* 34(4):744–754.
Moorer, J. A. 1974. "The Optimum Comb Method of Pitch Period Analysis of Continuous Digitized Speech." *IEEE Transactions on Acoustics, Speech, and Signal Processing* 22(5):330–338.
Moorer, J. A. 1977. "Signal Processing Aspects of Computer Music: A Survey." *Proceedings of the IEEE* 5(8):1108–1137.
Orfanidis, S. J. 1996. *Introduction to Signal Processing.* Upper Saddle River, New Jersey: Prentice Hall.
Pakarinen, J., V. Välimäki, and M. Karjalainen. 2005. "Physics-Based Methods for Modeling Nonlinear Vibrating Strings." *Acta Acustica United with Acustica* 91(2):312–325.
Pei, S.-C., and C.-C. Tseng. 1997. "IIR Multiple Notch Filter Design Based on All-Pass Filter." *IEEE Transactions on Circuits and Systems Part II: Analog and Digital Signal Processing* 44(2):133–136.
Pei, S.-C., and C.-C. Tseng. 1998. "A Comb Filter Design Using Fractional-Sample Delay." *IEEE Transactions on Circuits and Systems Part II: Analog and Digital Signal Processing* 45(6):649–653.
Pei, S.-C., and P.-H. Wang. 2004. "Closed-Form Design of All-Pass Fractional Delay Filters." *IEEE Signal Processing Letters* 11(10):788–791.
Rauhala, J., and V. Välimäki. 2006. "Tunable Dispersion Filter Design for Piano Synthesis." *IEEE Signal Processing Letters* 13(5):253–256.
Rocchesso, D., and F. Scalcon. 1996. "Accurate Dispersion Simulation for Piano Strings." *Proceedings of the 1996 Nordic Acoustical Meeting.* Helsinki, Finland: NAM, pp. 407–414.
Saramäki, T. 1985. "On the Design of Digital Filters as a Sum of Two All-Pass Filters." *IEEE Transactions on Circuits and Systems* 32(11):1191–1193.
Serra, X. 1989. "A System for Sound Analysis/Transformation/Synthesis Based on a Deterministic Plus Stochastic Decomposition." Ph.D. Thesis and Report No. STAN-M-58, Stanford University, Stanford, California.
Serra, X., and J. O. Smith. 1990. "Spectral Modeling Synthesis: A Sound Analysis/Synthesis Based on a Deterministic Plus Stochastic Decomposition." *Computer Music Journal* 14(4):12–24.
Steiglitz, K. 1996. *A Digital Signal Processing Primer.* Menlo Park, California: Addison-Wesley.
Thiran, J.-P. 1971. "Recursive Digital Filters with Maximally Flat Group Delay." *IEEE Transactions on Circuit Theory* 18(6):659–664.
Tolonen, T. 1999. "Methods for Separation of Harmonic Sound Sources Using Sinusoidal Modeling." *Proceedings of the 106th AES Convention.* New York: Audio Engineering Society, preprint 4958. Available online at lib.tkk.fi/Diss/2000/isbn9512251965/article7.pdf.
Tseng, C.-C., and S.-C. Pei. 2001. "Stable IIR Notch Filter Design with Optimal Pole Placement." *IEEE Transactions on Signal Processing* 49(11):2673–2681.
Välimäki, V. 2000. "Simple Design of Fractional Delay All-Pass Filters." *Proceedings of the 2000 European Signal Processing Conference.* Kessariani, Greece: EURASIP, vol. 4, pp. 1881–1884.
Välimäki, V., and A. Haghparast. 2007. "Fractional Delay Filter Design Based on Truncated Lagrange Interpolation." *IEEE Signal Processing Letters* 14(11):816–819.
Välimäki, V., M. Ilmoniemi, and M. Huotilainen. 2004. "Decomposition and Modification of Musical Instrument Sounds Using a Fractional Delay All-Pass Filter." *Proceedings of the 2004 Nordic Signal Processing Symposium.* Espoo, Finland: IEEE Finland Section, pp. 208–211.
Välimäki, V., and T. I. Laakso. 2001. "Fractional Delay Filters: Design and Applications." In F. Marvasti, ed. *Nonuniform Sampling: Theory and Practice.* New York: Kluwer Academic/Plenum Publishers, pp. 835–895.
Välimäki, V., H.-M. Lehtonen, and T. I. Laakso. 2007. "Musical Signal Analysis Using Fractional-Delay Inverse Comb Filters." *Proceedings of the International Conference on Digital Audio Effects.* Bordeaux, France: LaBRI, University of Bordeaux 1, pp. 261–268.